Verbundvorhaben

# Semantic Transfer in Verbmobil

Ann Copestake

Universität Stuttgart and CSLI

Report 93
August 1995

August 1995

Ann Copestake

Institut für Maschinelle Sprachverarbeitung
Universität Stuttgart
Azenbergstrasse 12
70174 Stuttgart

Tel.: (0711) 121 - 1363
Fax: (0711) 121 - 1366

CSLI Center for the Study of Language and Information
Ventura Hall, Stanford University
Stanford CA 94305-4115

Tel.: 001 415-725 2312
Fax: 001 415-723 0758
e-mail: `aac@csli.stanford.edu`

**Gehört zum Antragsabschnitt:** 12 Transfer

Die vorliegende Arbeit wurde im Rahmen des Verbundvorhabens Verbmobil vom Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF) unter dem Förderkennzeichen 01 IV 101 G gefördert. Die Verantwortung für den Inhalt dieser Arbeit liegt bei der Autorin.

## Abstract

This paper is a detailed discussion of semantic transfer in the context of the Verbmobil Machine Translation project. The use of semantic transfer as a translation mechanism is introduced and justified by comparison with alternative approaches. Some criteria for evaluation of transfer frameworks are discussed and a comparison is made of three different approaches to the representation of translation rules or equivalences. This is followed by a discussion of control of application of transfer rules and interaction with a domain description and inference component.

## 1 Introduction

Verbmobil is a spoken-dialogue MT system, which takes German language input and produces English output. In the near future it will also operate with Japanese input and English output. The system currently assumes a small vocabulary and is restricted to the task of arranging meetings but intended to be extensible (in the medium to long term) to less restricted topics. Because of this ambitious long-term goal, components of Verbmobil must be well motivated in design, to attempt to avoid, or at least delay, the problems in scaling up which have so consistently hit previous natural language systems. MT systems, in particular, appear to be extremely difficult to scale up, and the aim of this paper is to do a theoretical study to investigate the potential problems.

We describe a range of approaches to translation and discuss their suitability for Verbmobil, taking account of both the medium and the long term goals. We will attempt for the most part to abstract away from existing implementations, but to keep the discussion grounded by considering examples taken from the task-oriented dialogues collected for Verbmobil. However, as far as possible, we will attempt to generalise from these examples to classes of problems that might arise, both within the chosen domain and with a less restricted task. The intent of this paper is to follow on from Kay *et al* (1994) and to make a more detailed and concrete evaluation of approaches to MT in the light of experience already gained on Verbmobil (and on other projects). We will assume the use of HPSG for the syntactic representation (Pollard and Sag, 1994), DRT for the semantics (Kamp and Reyle, 1993) and an underlying unification/constraint-based implementation framework. We will only consider German to English translation specifically here, due to lack of sufficient Japanese data for the Verbmobil domain.

This paper is organised as follows: in §2 we briefly review existing approaches to translation. Following Kay *et al* (1994) we will argue that the most appropriate style of approach is one in which analysis produces a (underspecified) semantic representation, which is not expected to be identical to the semantic representation for the target language. We will refer to this as semantic transfer, but we construe this rather broadly. As we will explain, we include systems which make some reference to syntactic structure as well as semantics and also approaches which make use of a (partial) interlingua but do not insist that a single representation exists for source and target utterances. We also include systems based on semantically driven lexicalist transfer. We will diverge somewhat from Kay *et al*'s suggestions in that we assume a rather greater role for transfer and less reliance on interlingual representation. In §3 we lay the groundwork for a detailed comparison of three particular approaches to stating transfer rules. In §4, §5 and §6 we describe the three approaches in detail, by illustrating how they work on some sentences from the Verbmobil dialogues. We use other examples to examine different facets of translation and expose some limitations in the approaches. In §7, we discuss the possibilities for interface with a domain model and inference engine with respect to these approaches, and outline ways in which the application of transfer rules may be controlled. Finally, §8 summarises the most important findings.

## 2 A classification of approaches to MT

The following description of MT approaches is intended more as an indication of the nomenclature we will use than as a full discussion, for which see e.g. Hutchins and Somers (1992). Kay *et al* (1994) discuss these approaches in the Verbmobil context. The references to various systems given below are not intended to represent an absolute categorisation, since the boundaries between these approaches are not sharp. In particular, inclusion of a system in a particular category does not necessarily imply that it suffers from any problems which are identified for that approach in general, since some deficiencies may be alleviated by adopting techniques from other approaches.

**Direct transfer** (e.g. SYSTRAN, though this has further refinements; nearly all portable and PC based systems). Very limited processing of the source language (SL) string is followed by transfer to a target language (TL) string. Some morphological analysis is usual but no parsing or semantic analysis takes place. The limitations are obvious: vast numbers of transfer rules are required for anything other than very limited and repetitive text, the only disambiguation is based on collocations, the output is likely to be ungrammatical. SYSTRAN itself is not as limited as this description would suggest, since the current version incorporates some deeper analysis.

**Syntactic transfer** (e.g. the METAL system, Slocum *et al*, 1987). Transfer is carried out on syntactic trees. It therefore requires that a syntactic analysis of the input be possible, which requires a significant amount of processing, and a large grammar and lexicon (for a broad coverage system). The transfer component will have fewer rules than in the direct transfer method, but the rules are more difficult to write, since they must refer to abstract syntactic categories rather than surface strings. The degree of similarity between source and target representations depends on the syntactic analysis. However, a syntactic transfer approach makes little sense when using HPSG, since the phrase structure trees corresponding to the source and target languages will often be very different, and the semantics is constructed in parallel with the syntax.

**Unification-based syntactic/semantic transfer** (e.g. Zajac (1989), Kaplan *et al* (1989), ELU — Estival *et al* (1990), MiMo — Arnold and Sadler (1990)). Unification-based systems can treat transfer rules as being similar to grammar rules, which map between feature structures. Such systems have many of the good points of unification-based grammars: they are (to a large extent) declarative and can therefore accommodate different parsers and generators, and they can be bidirectional. The systems mentioned above rely mostly on stating semantic relationships: it is not entirely clear whether the capability to express syntactic mapping is ever really necessary, because where reference to syntax is made it could be argued that this is only required because of deficiencies in the monolingual semantic analysis. The complexity of the transfer rules depends on the syntax of the semantic representation. We will discuss this in more detail below.

**'Pure' semantic transfer** (e.g. BCI — Alshawi *et al*, 1991). Transfer is carried out between semantic representations. The distinction between this and the previous class is not clear-cut, however, as some syntactic information may also be accessed. The BCI system, for example, uses information about whether a sentence is active or passive in transfer. The use of the term 'transfer' implies that the mapping is non-trivial — i.e. that the semantic representations of the source and target language are not identical and may use different predicate names. However, there is no clear distinction between semantic transfer and an interlingua system that does not insist that source and target semantic representations are identical, since most interlingual systems have some language-specific predicates and most semantic transfer approaches make use of some interlingual concepts (to represent tense, for example). We will discuss pure semantic transfer in more detail below.

**Interlingua** (e.g. Dorr (1994), Emele *et al* (1992), Landsbergen (1987)). In the prototypical ('naive') interlingua approach, an identical representation is always used for source and target language utterances. One advantage which is generally claimed is that, while transfer between $n$ languages requires that $(n-1)^2$ sets of transfer rules be built, an interlingua approach only requires $n$ sets (one for each language) mapping to the common representation. However, this argument implies that an interlingua is not an asset in translation between two languages (one transfer rule-set vs. two language-to-interlingua rule-sets) or for the Verbmobil scenario (two unidirectional transfer rule-sets vs. three language-to-interlingua sets). Below we discuss some reasons to doubt that this pure form of interlingual approach can be achieved in any case.

**Lexicalist transfer** (Shake-and-Bake — Whitelock (e.g. 1992) and variants: e.g. Beaven (1992), Copestake *et al* (1995), Trujillo (1995)). This approach relies on the use of a lexicalist unification-based grammar formalism. Translation equivalences are stated by relating the semantics of source and target language lexical signs. It has been claimed that this combines the advantage of simplicity of description of transfer rules found in the direct approach, with some of the advantages of semantic or syntactic/semantic transfer. This is also discussed in more detail below.

**Knowledge-based machine translation** (e.g. Nirenburg and Raskin, 1988). This approach emphasises reliance on domain knowledge in carrying out translation. It is based on extensive (and time-consuming) domain modelling. It has been used for tasks such as translation of technical manuals written in a subset of a natural language which is constrained with respect to vocabulary and grammar. The approach is often claimed to be an interlingual one, but the language-independent representation is domain-specific, and some transfer rules may be used for convenience. The problem with this approach for Verbmobil is its domain and task specificity — while it is reasonable to suggest that technical manuals refer to a strictly circumscribed domain, it is much more difficult to defend this position with respect to the Verbmobil dialogues and the meeting arrangement task. It is possible to regard meeting arrangement in an abstract way as steps in a process to narrow down times and locations to achieve compatibility. However, it is clear that a lot is going on in the Verbmobil dialogues which is non-essential and unpredictable from this viewpoint, but is an inherent part of human communication. For example, excuses for not being able to meet at a particular time cannot be treated as part of a domain model, because the possible excuses are unbounded. Therefore, we regard domain knowledge as an important component in a translation system, but do not believe that basing the system completely on the domain is a good long-term strategy.

**Statistical approaches** (e.g. Brown *et al*, 1990). These are only relevant to a situation where a very large bilingual corpus exists, and only work well in situations where very local context gives enough information to translate a word. For these reasons, such approaches appear inapplicable as the main approach to translation in Verbmobil and thus we do not consider them further here (see Kay *et al* for further details).[1]

[1] Statistical approaches may have a role to play in control of translation, however, as mentioned in §7.3.

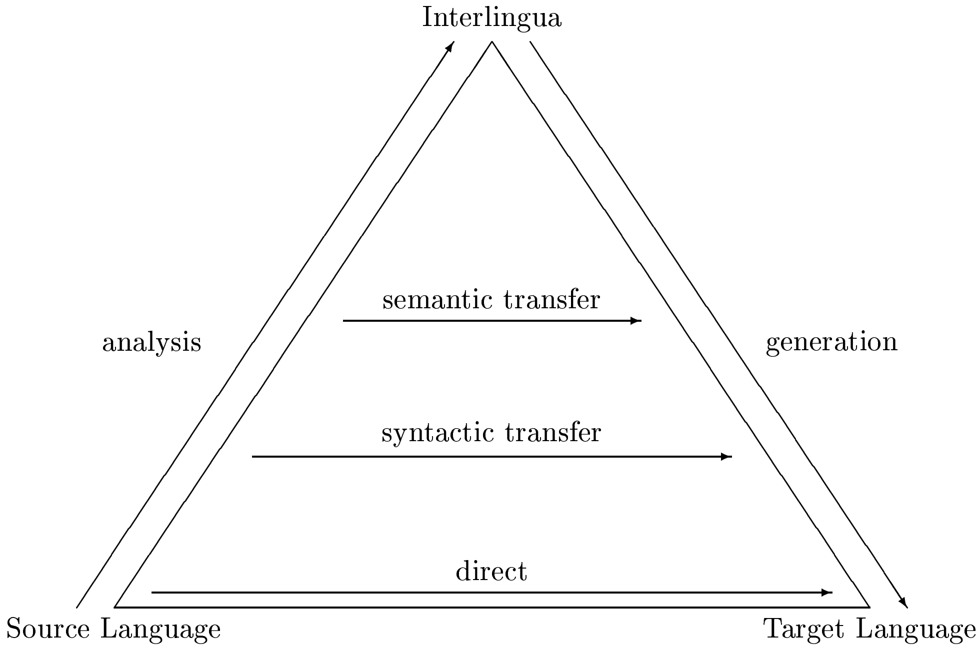


Figure 1: Vauquois triangle

### 2.1 Triangles, polygons and naive interlingual approaches

The classic way of describing the tradeoff between some of these approaches to translation is the Vauquois triangle reproduced in Figure 1 (adapted from Hutchins and Somers, 1992). The vertical dimension is intended to represent the degree of difficulty in analysing the source to the chosen level and in generating the target string from the corresponding target language representation. The horizontal dimension is intended to represent the degree of difficulty in carrying out transfer (i.e. converting the analysed source representation to the target representation). On this picture, the direct approach takes the least effort for analysis and generation, but actual transfer is most difficult. In contrast, the apex of the triangle represents a naive interlingual system in which a single representation is constructed for the SL from which the TL string can be generated. This makes transfer unnecessary but increases the difficulty of analysis and generation. Syntactic transfer and semantic transfer are seen as being on intermediate points in the scale.

Given the current state of the art in natural language processing, we believe that a more accurate picture is that shown in Figure 2. This diagram is not intended to be taken too seriously, but expresses several divergences from Vauquois'. The most important of these is skepticism about the possibility of a single language-neutral sentence representation. If there is a possible language-neutral representation, it is at the level of the utterance, rather than the sentence: i.e. context must be taken into account. But current NLP systems have only made limited progress towards this sort of analysis. MT is feasible because the speaker and the hearer to a large extent share the same world knowledge and context model, so the translation system can avoid doing some of the hard work, if the translated sentence makes the "correct" set of inferences available to the hearer. For example, a literal translation of *the window is open* into German would probably be interpreted by the hearer as a request to close it, in a context in which the English utterance would be interpreted in that way. But an MT system would be unable to accurately decide whether the sentence should be interpreted as a request, because so much world knowledge could influence this:

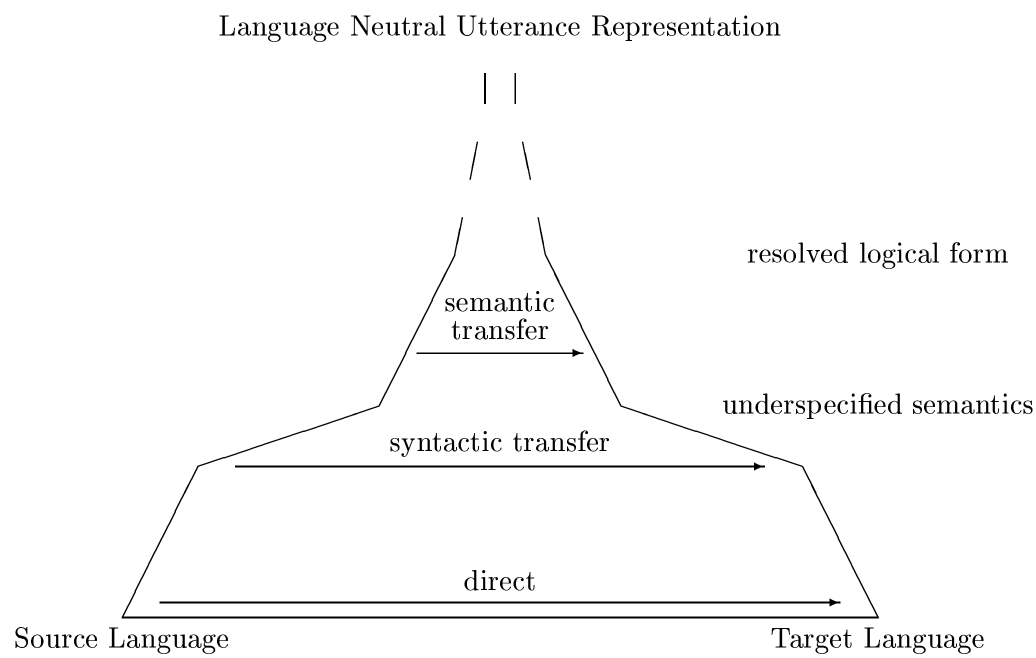


Figure 2: Modified Vauquois 'triangle'

room temperature, noise level, odours etc, etc.

Kay *et al* (1994) discuss the difficulties with the naive interlingual approach in some detail, so here we simply summarise their argument, which is basically that some sentences cannot be idiomatically translated in a way that completely preserves meaning. One example given by Kay *et al* (1994:26) is:

(1) English: He asked where he should stand.
French: Il a demandé où il devait se mettre.
Lit: He asked where he should put himself.

In this case, the French example is less specific than the English source, but a more literal translation would be awkward.

An example from the Verbmobil dialogues, which illustrates the same point, is given in (2):

(2) German: Da habe ich immer schon Feierabend
Lit: Then I have always already time-after-work
English: I have (usually) finished work by then

There are two sources of difficulty in finding a good translation: the noun *Feierabend* does not have a straightforward equivalent in English (*knocking off time* is the closest, but is point-like while *Feierabend* refers to an interval, and *knocking off time* is mostly used for factory work) and it is difficult to find a way to convey the same force of the habitual (which is less strong than the literal translation by *always* would suggest). In many contexts, the English sentence given would be equivalent to the German one, but it is more general, since *finish work* could refer to work other than an actual job, or to a time period other than the whole day (e.g. that sentence could be used in a different context to refer to the time a lunchbreak started). It also illustrates a general problem which can occur if a noun is not translated as a noun: subsequent pronominal reference is possible to *Feierabend* in the German discourse, but not in the English:

(3) a Speaker A: Da habe ich immer schon Feierabend
Speaker B: Oh, in meiner Firma ist er später.

b Speaker A: I have (usually) finished work by then.
Speaker B: ? Oh, it's later at my company.

It is therefore clear that the sentences in (2) cannot have the same DRS.

There are other cases of lack of translation equivalence which involve cultural differences. For example, in the Langenscheidts Taschenwörterbuch, *Assessor* is given a definition in English rather than a translation: *civil servant who has completed his/her second state exam.* Such examples are frequent in bilingual dictionaries and cause considerable problems when trying to construct a transfer lexicon automatically (Copestake *et al*, 1995). However, these are not strictly speaking translation problems, since German language speakers from another country would have the same difficulty in understanding *Assessor* as English speakers would. One of the few cases of a clear cultural difference that shows up in the Verbmobil dialogues is the occasional reference to *Kalenderwoche* (weeks of the year):

(4) German: Im März höchstens die zehnte Kalenderwoche aber nur ungern.
Lit: In March at best the tenth calendar week but only reluctantly.

The concept could be expressed precisely in English, of course, but only in a very long-winded way, because there is a convention as to when the first *Kalenderwoche* begins which has to be explicitly expressed. It is not a usual way of referring to dates in either the UK or the USA and even in Germany is only usual among businessmen. These cases are clearly different from the *Feierabend* example, since that concept is universal in Western industrialised countries but is just not lexicalised in the same way in English.

It may always be possible to express the same truth conditional meaning in another language by means of some complex paraphrase, but this is irrelevant to the problem of translation, since it is unacceptable to produce verbose TL strings given a simple SL input. This is particularly true of spoken dialogue systems. A translation which was meaning preserving but verbose would fail to serve the communicative function of the utterance.

Thus, as we said at the beginning of the section, if there is a language-neutral representation, it is at a different level: one in which the communicative intentions of the speaker are modelled. [2] However, as the dashed lines in Figure 2 are intended to convey, we have relatively little knowledge of how to analyse a discourse to achieve a model of the speaker's communicative intentions, compared to our knowledge of syntactic and semantic analysis. As the gradient of the lines is intended to suggest, it is certainly difficult and computationally expensive, even if it is achievable in principle with a restricted domain. Furthermore, the current assumption about the use of Verbmobil is that the machine will only be required to translate parts of a dialogue and that, because of the limitations of speech recognition technology, it will not be able to monitor portions of the dialogue it is not required to translate. Even without this limitation, it will, of course, miss any non-verbal communication (gestures etc) which the participants in the dialogue make. It is therefore unrealistic to assume that Verbmobil could maintain a sufficiently accurate discourse model for a language-neutral utterance representation to be possible.

This is not to deny that context is important in translation. There are cases where it is impossible to translate without having some model of the discourse state and the speaker's intentions. For example, within the Verbmobil dialogues, there are two quite different possibilities for translation

[2] This does not apply to systems where the application is much more circumscribed and conventionalised, such as translation of weather reports and of the restricted language technical manuals mentioned earlier. Here a language-independent representation may be possible, though highly task-specific.

of (5a) into English, shown in (5b) and (5c):

(5) a Geht es bei Ihnen?

b Does that suit you?

c How about at your place?

We will therefore argue for a model in which domain and discourse information can be accessed by transfer. Initially this access would be quite limited, but as techniques for discourse analysis and speech recognition improve, we would expect transfer to rely more heavily on context.

There are also straightforward engineering reasons for rejecting approaches which require identity of source and target language representation, because they require that the source and target language grammars are developed together, in order to ensure the meaning representations are identical (or at least compatible). It is quite difficult enough to build a robust grammar without this constraint. For Verbmobil it is practically impossible to ensure that English and German grammars produce identical representations, since the grammars are being developed by different sites. Building extra modules, to convert the representations produced by the grammar into a unique interlingual representation, is also unattractive, since it would essentially require three transfer components instead of two (German, English and Japanese to interlingua, as opposed to German-to-English and Japanese-to-English).

We therefore reject a model of translation in which it is necessary to produce a single representation, in favour of one in which the source and target representations may differ. The assumption is that translation can be accomplished in many cases without attempting a full calculation of communicative intention, and thus that transfer at the level of semantic representation of the sentence is the most sensible option, provided that the transfer component can interact with other modules in the cases where this fails. Thus, for (5a), the transfer component might query some other module which would not have to maintain a full discourse model, but would have to discriminate between a state of the dialogue where agreement on a prior proposal was sought (where (5b) would be appropriate) and another state where a location for the meeting was being suggested (where (5c) would be required).

### 2.2 Underspecification

The second main point that Figure 2 is intended to convey is that it is possible to have underspecified semantic representations. This has been discussed by Alshawi *et al* (1991) with respect to the decision to use QLF (quasi-logical form) as the transfer level in the SRI BCI system, rather than the fully instantiated logical form. The general issue is that disambiguation can involve an arbitrary amount of reasoning on real-world knowledge, which is highly problematic in an implementation. If the ambiguity is preserved, however, the pragmatically plausible reading will be the same in the target language as in the source, and the hearer will normally get the right reading without conscious effort. Several kinds of underspecification are potentially relevant:

**Attachment ambiguities** PP-attachment ambiguities are notorious problems in analysing English, since the number of possible parses for a sequence of PPs increases according to the Catalan series, and disambiguation can involve arbitrarily complex reasoning. Sequences of four or more PPs are not uncommon. Similar remarks apply to German. In German to English translation, it is often possible to straightforwardly translate PPs as PPs maintaining the relative ordering, and in this case the preferred reading will usually remain the same.

(6) a Jeder Nachmittag ist frei in meinem Terminkalender ab vierzehn Uhr.

b Every afternoon is free in my appointment book after two o'clock.

Similar remarks apply to other attachment ambiguities, such as those involving coordination:

(7) a Meinen Sie Donnerstag den 8. oder Donnerstag den 15. Juli?

b Do you mean Thursday the eighth or Thursday the fifteenth of July?

It is pragmatically most plausible that *Juli* has wide scope here, but *den 8* could refer to April 8th. However, it is unnecessary to attempt this disambiguation, because the hearer is almost certain to get the right reading, because he or she has the appropriate context and world knowledge. It should be noted that this requires an interface with syntax which allows attachment to remain underspecified.

**Quantifier scope** In contrast, quantifier scope ambiguities do not correspond to different syntactic structures, although syntax does play a role in determining which readings are possible. In (8) is a (made-up) example, where dual readings are plausible (at least for English).

(8) a Jeder Student in diesem Studiengang muss vier Programmiersprachen lernen.

b Every student on this course must learn four programming languages.

Conventional grammars give some scopes which are not at all plausible to human readers but excluding implausible readings is not straightforward. Thus excluding the unwanted reading from (9) (which is taken from the CMU collection of English dialogues for the Verbmobil domain) is not totally trivial:

(9) I have some time almost every day.

We will discuss one technique for the representation of underspecified scope in §6, but here we should point out that quantifier scope ambiguities may not be too serious for Verbmobil, since German word order usually disambiguates scope and potential ambiguities are relatively rare in the Verbmobil domain: (9) is one of the few examples found with an explicit universal, for instance.

**Pronouns** Pronoun resolution is another area in which current NLP systems cannot achieve perfect results, although the algorithms are generally better than those available for PP-attachment. For German-to-English translation, pronouns can be left partially resolved: that is, it is necessary to know whether a third person pronoun refers to a human or not (to avoid *er* being translated as *him* when it actually refers to a *Termin*, for example) but it is unnecessary to know which human or object it refers to. Thus, it is unnecessary to resolve *es* in (10a) in order to translate it to (10b):

(10) a Zur Not können wir es auf den ersten verlegen.

b If necessary, we could move it to the first.

In fact, pronouns are mostly first and second person in the collected dialogues.

**Compounds** Attachment ambiguities are a problem when analysing compounds with more than two constituents. But, quite apart from this, it is desirable to be able to underspecify the relationship between parts of a compound, because determining the correct relationship is potentially arbitrarily complex. How serious this is for Verbmobil in its current state depends on whether the vocabulary limitation is taken to apply to compounds (e.g. if *Termin* and *Zahnarzt* are in the vocabulary, is there a requirement to recognise *Zahnarzt-Termin* even if it is not listed explicitly?). If productive compounds can occur they should be translated as compounds in English where possible. For example:

| | |
|---|---|
| *Mittags-Termin* | *midday meeting* |
| *Berlin-Videos* | *Berlin videos* |
| *Fasanengarten* | *pheasant garden* |

However, this strategy will not work in all cases and it may be necessary to resolve the relationship:

*Terminvorschlag* *proposal for a meeting*

This particular example is discussed in more detail in §7.2.

**Lexical ambiguity** Ambiguity between homonyms, such as *bank*:'side of a river' as opposed to *bank*:'financial institution' will nearly always have to be resolved in translation, since pairs of languages will only share homonyms occasionally. But regular polysemy is often language-independent (in Apresjan's (1973) collection of examples of regular polysemy, about two-thirds of the patterns had counterparts in both Russian and English). For example, *Zeitung* and *newspaper* show the same physical object/organisation polysemy. More relevant to Verbmobil are the unestablished cases of polysemy: for example (11a) and (11b) both have the same metonymic interpretation which is given in (11c):

(11) a Nach Stuttgart treffen wir uns in Saarbrücken.
b After Stuttgart we will meet in Saarbrücken.
c After the meeting in Stuttgart we will meet in Saarbrücken.

For further discussion of regular polysemy see Copestake and Briscoe (1995) and references therein.

### 2.3 Semantic transfer

Figure 2 expresses a considerable bias in favour of (underspecified) semantic transfer rather than syntactic transfer, in that constructing an underspecified semantic representation involves relatively little effort for its return in narrowing the gap between representations. Again this has been discussed elsewhere, including Kay *et al* (1994), Alshawi *et al* (1991). Here we simply make two points: the first is that in HPSG syntactic and semantic processing are very directly interrelated, thus in computational terms the extra overhead involved in building a semantic representation may not be great. How much effort is involved will depend on the form of the semantic representation chosen and how directly it relates to the syntax. Secondly, the closeness in representation between source and target languages that is achieved will also depend on the semantic representation. If very different representations are chosen, it is possible that the sides of the 'triangle' could actually diverge rather than converge. Again it is often the form of the representation which is crucial here, and not the model theory.

Our conclusion in favour of semantic transfer is broadly in line with Kay *et al*'s suggestions for Verbmobil, though we have used some slightly different terminology. We also accept their proposal for translation by negotiation, but in this paper we will concentrate on the declarative aspects of some existing formalisms. We have no firm proposals to make about how to achieve the (highly desirable) aim of very fluid interaction between components envisaged by Kay, though we will note some potential limitations of existing formalisms in this respect. One difference, however, is that Kay's model does not include any notion of transfer rules or equivalences and relies on the use of an interlingua. The generator is required to approximate the source language representation as closely as possible, but it is not clear how this is to be achieved. We think that it is necessary to give more guidance to the generator, and the best way to do this is to take advantage of knowledge about divergences between languages and to write language-pair specific rules to transfer between source and target languages.

We also feel uneasy about the use of an interlingua, because it implies strict identity between concepts in different languages which is not necessarily plausible, except for terms which have externally imposed definitions. For example, it is reasonable to assume that literal usages of German *Elefant* and English *elephant* might have identical denotations, because they both refer to the same two species of mammal, but this is only clear for (some) natural kind terms. Sometimes differences in meaning are very subtle: *geben* is normally translated as *give* when it refers to a transfer of possession, but it is sometimes used in contexts where *lend* would be a much more appropriate translation in English[3]. Such mismatches are not going to appear in bilingual dictionaries, and it is probable that they will not become apparent until a particular context is encountered that displays the difference.

If the *geben/give* mismatch was discovered after we had associated an interlingual term GEBEN with the entries for (particular senses of) *give* and *geben* in the monolingual lexicons, we would be forced to change those entries. One way of doing this would be to invent another sense for *geben* which would correspond to the same interlingual predicate as *lend*. On monolingual grounds, however, this does not seem to be justified, and it is unrevealing — we would need to be able to disambiguate between the senses in order to translate, so we would need an explicit statement of the differences between them. As an alternative, we could make the semantics of *give* more complex, as shown below (this analysis is given for the sake of concreteness, but is not fully adequate):

(12) geben: $\mathrm{GEBEN}(e, x, y, z)$
give: $\mathrm{GEBEN}(e, x, y, z) \wedge NOT(BELIEVE(\text{speaker}, \exists e'[GEBEN(e', y, x, z) \wedge AFTER(e, e')]))$

This is intended to express the idea that *geben* might be neutral with respect to future transfers but that *give* implies permanence. In principle, we might want to explicitly state information like this about the meaning of *give*, but there is no monolingual reason to put it in the lexicon rather than treat it via inference rules. There is no obvious bound on the complexity that might be involved, and since this strategy involves using other interlingual predicates in the lexical entries, changes could have ramifications throughout the lexicon (e.g. an English native speaker could have used GEBEN in other entries assuming it was equivalent to *give*).

Extrapolating from the example given in Kay *et al* (1994), we assume that, in the translation by negotiation approach, the generator would be passed a logical form containing $\mathrm{GEBEN}(e, x, y, z)$ and then discover that the 'nearest' lexical entry was *give* but that this carried some extra implications.[4] The generator would then check with the negotiator to see if these extra conditions held. But a transfer-based approach could provide greater modularity. There would be language-specific predicates for *geben* and *give* which would be related by a simple transfer rule (the connective $>$ is intended to suggest default implication):

geben: $geben'(e, x, y, z)$
give: $give'(e, x, y, z)$
$geben'(e, x, y, z) > give'(e, x, y, z)$

The exceptional cases could be represented as follows:

[3] Thus *Du gibst mir das Buch* could be said when the speaker had borrowed the book two days ago but *You gave me that book* is judged to be inappropriate by most English speakers in this situation. This use is not equivalent to examples where *give* is used instead of *pass*, (e.g. *Give me your plates so I can put them in the dishwasher*) which is also possible in German. It is more similar to the use in *I gave him my watch while I went swimming*, but this is more restricted in English than the German use of *geben* is, since it requires a very explicit context to limit the duration of the transfer. Thanks to Susanne Riehemann for a detailed discussion of this.

[4] There are severe difficulties in formalising the concept of 'nearness', but we will ignore this issue here.

$geben'(e,x,y,z) > lend'(e,x,y,z)$
$BELIEVE(\text{speaker}, \exists e'[geben'(e',y,x,z) \wedge AFTER(e,e')])$
Here the second line is a condition on the applicability of the rule. Such transfer rules have a nonmonotonic interpretation, since they may fail. The more specific rule should apply in preference to the more general one if its conditions are met. However, an equivalent monotonic formalisation is possible, though less concise. This formulation is more-or-less equivalent to the interlingual solution described above in its effect, but the advantage of locating this information in a transfer component is that the monolingual lexicons are less complex and less subject to changes which are unjustified by monolingual criteria.

Thus our basic argument for semantic transfer is that knowledge of word meaning is very imperfect. Minor discrepancies between the meanings of the source and target language words will normally not even be noticed, but occasionally contexts may arise where a subtle distinction becomes important. If we try and make predicates language-independent, a discovery of a mismatch could require extensive changes to the monolingual lexicons. On the other hand, the use of semantic transfer rules can be seen as an admission that a correspondence may not be exact: we are not claiming that the denotations of the predicates are identical, just that they are close enough for the contexts we know about. This level of indirection makes it easier to adjust for discrepancies as they become known.[5]

Transfer approaches are powerful enough to relate arbitrary expressions in the source and target language semantics and because of this it is easy to end up with an unmanageable set of transfer rules. Transfer rules are often used to attempt to fix problems that are really caused by idiosyncrasies in the target language, rather than a difference in meaning between source and target language predicates. There is also a temptation to try and build rules that give a natural sounding translation on a particular example, even though this may not generalise. Later in this paper, in particular in §7, we will give some examples where we think it is best to assume that the monolingual components should do some of the work of lexical choice and of blocking bad translations.

### 2.4 Logical form equivalence and flat semantics

Figure 1 and Figure 2 are both somewhat misleading in the symmetry they imply between analysis and generation. There are well-known theoretical problems in ensuring that a grammar can generate from a particular semantic representation which been discussed in the context of generation by Shieber (1993) and in machine translation by Landsbergen (1987) and Whitelock (1992), among others. One difficulty is the problem of logical form equivalence: even though the grammar may accept a logical form (LF) logically equivalent to a particular LF which is input to the generator, there is no guarantee that it will generate from that syntactic form of the input LF. To take a trivial example, an English grammar might naturally produce the logical form in (13a) from *fierce black cat*, while a straightforward transfer from the natural Spanish representation of *gato negro y feroz* shown in (13b) would produce the LF in (13c), which the English grammar probably would not allow:

(13) a $\lambda x[\text{fierce}(x) \wedge (\text{black}(x) \wedge \text{cat}(x))]$
b $\lambda x[\text{gato}(x) \wedge (\text{negro}(x) \wedge \text{feroz}(x))]$
c $\lambda x[\text{cat}(x) \wedge (\text{black}(x) \wedge \text{fierce}(x))]$

[5] This may imply that semantic transfer rules could only be satisfactorily formalised in a nonmonotonic system, but we will not discuss that issue further here.

One possible solution to this problem would be for the generator to try all logically equivalent forms. Unfortunately this is not practicable in general since the logical form equivalence problem is undecidable even for first order predicate calculus (Shieber 1993).

Both Landsbergen's approach and the Shake-and-Bake framework for MT were developed in response to this problem. We have already suggested that the difficulty of developing grammars which produce identical representations makes the naive interlingual approach unsuitable for Verbmobil. Landsbergen's approach is even more restricting, since it ensures identity of the interlingual representation by insisting that the grammars for source and target language be isomorphic, which is practically extremely difficult to achieve even for closely related languages such as English and German (see e.g. Beaven (1992) for the difficulties of developing isomorphic grammars). If, as we suggested above, translation does not necessarily (or even usually) involve preservation of denotation, this is clearly not the best approach. We will discuss Shake-and-Bake style approaches in §5.

A related issue, which is also addressed by Shake-and-Bake, is the statement of phrasal transfer equivalences where material may be interpolated between lexical items. For example, assuming the representation of adjectives shown above, we cannot simply state the following transfer rule for the equivalence English:*young bull* ≈ Spanish:*novillo*:

$$\lambda x[\mathrm{young}(x) \wedge \mathrm{bull}(x))] \leftrightarrow \lambda x[\mathrm{novillo}(x)]$$

Interpolated material is possible. For example, *young black bull* could have the following logical form:

$$\lambda x[\mathrm{young}(x) \wedge (\mathrm{negro}(x) \wedge \mathrm{bull}(x))]$$

Either a more complex transfer rule has to be written which allows for the intervening structure, or the transfer component has to be able to rearrange the logical form, so that the rule does apply as written, which implies that the transfer component needs knowledge of the logical properties of the connectives, and is computationally expensive. In general, the greater the structural complexity of the representation, the worse the problem of constructing a LF acceptable to the target grammar becomes. Copestake *et al* (1995) discuss these issues in greater detail, and suggest an approach to a 'flat' or minimally recursive semantic representation which alleviates some of the problems. We discuss this further below, in §6.

The overall conclusion of this section is that some form of semantic transfer approach is most appropriate for Verbmobil, since an interlingual system would require much better computational treatments of pragmatics and lexical semantics than those currently available, given that the Verbmobil application is not tightly circumscribed. Transfer on an underspecified semantic representation is viable because it exploits the hearer's world knowledge to make up for deficiencies in automatic ambiguity resolution and pragmatic processing, and utilises information about divergences between specific languages to avoid the need to provide a detailed lexical semantic analysis for every word. However, a transfer approach certainly does not obviate the need for pragmatics and lexical semantics, since the support of a discourse analysis component is often required to resolve ambiguities, and appropriate monolingual lexical representations can greatly reduce the number and complexity of the transfer rules. We will attempt to illustrate this with some of the examples that follow in the rest of this paper.

# 3 Comparison of transfer formalisms

Having identified semantic transfer as the most appropriate approach for Verbmobil, in this section we make some general remarks about transfer formalisms which will act as background to the detailed discussion in the subsequent sections.

## 3.1 Criteria for comparison

The list of criteria below is intended as a general informal aid for comparison. It is not complete; categories are interrelated and to some extent arbitrary, and there is some overlap between them. There are tradeoffs between desirable features, and the weight given to them will vary with the application. We have put an emphasis here on features such as ease of acquisition, modularity, and independence of rules, because we are concentrating on the long-term development of MT systems such as Verbmobil and for this it essential that transfer rules be as easy to construct and maintain as possible.

**Coverage** Are there some phenomena which cannot be handled or cannot be dealt with elegantly? How important are they?

For theoretical purposes, this should be the main criterion. A transfer formalism should allow for any possible translation, just as a grammatical formalism should be capable of dealing with any sentence of the language. But, as with grammar formalisms, there are many caveats. All systems can trivially deal with any specific translation by simply stipulating the input/output pairing for that particular case. The is clearly appropriate for fixed conventional phrases such as *auf Wiedersehen*, but it means that the criterion of coverage has to take into consideration generality and elegance. However, for a practical system, it may not be worth expending effort to treat some very rare construction.

**Modularity** What assumptions are being made about the source and target language grammars in the transfer component?

As we mentioned above, some approaches to transfer essentially require that the source and target grammars are developed in parallel. This is undesirable in principle, because it does not allow reuse of grammars developed for other purposes, and also creates problems when extra languages are added. There is also the practical difficulty of managing grammar development under these conditions, which would probably be insuperable for Verbmobil since the English and German grammars are being developed on different sites.

None of the approaches discussed in detail here require isomorphism between grammars or identity of LF for target and source sentences, but, at least implicitly, they place some requirements for commonality between the source and target language semantics. It is an advantage to minimise this if it distorts the grammar so that the analysis becomes unnatural from a monolingual viewpoint.

**Independence** Can transfer rules be written independently of each other so that they nevertheless interact correctly?

Although this is not a topic that is discussed at length in the MT literature, it appears that ensuring that transfer rules interact correctly with each other and with the monolingual grammars has been one of the main problems in attempting to build robust MT systems and to scale-up systems to have reasonable coverage. Examples of where this might be problematic will be given later in the paper.

**Underspecification** Is transfer between underspecified representations possible?

The utility of this was discussed in §2.2.

**Generalisation** Can useful generalisations be made to reduce redundancy in transfer statements?

We will give examples later in the paper where generalisations about classes of transfer rules are possible.

**Interaction with generation** How flexible is the interaction with generation?

Some approaches to transfer, in particular Shake-and-Bake, cannot be used with conventional generators. Other approaches, which produce a logical form as the output of transfer, are more flexible. The input to generation can be underspecified in various ways. Some lexical choice is probably better left to the target grammar and the generator — prepositions are a fairly clear case where it is might be possible to have a complete interlingua, but where the choice of preposition is very often idiosyncratic and lexically governed. For example, an adequate English grammar should make the correct distinction between *in* and *on*, in order to generate *in the afternoon*, *on Thursday afternoon* but not **on the afternoon*, **in Thursday afternoon*. It is therefore unnecessary to attempt to do this in transfer. The prepositions *in* and *on* have the same meaning in these examples, so the selection of the actual preposition should be made by the monolingual system. Note the difference between this and the example of *geben*≈*give/lend* discussed in §2.3. Transfer should be responsible for the choice of *lend* rather than *give* in the context discussed, because the sentence *You gave me that book* could not be excluded by the target grammar.

**Interaction with pragmatics and the domain model** How would an inference engine interact with transfer?

We will discuss this in very general terms in §7, showing examples where inference would be needed to get the correct (or best) translation. The simplest method of interaction is to assume that the inference component is responsible for selecting translations or transfer rules from a range of possibilities provided by the transfer component proper. However there could be advantages if the formalism supports a tighter interaction.

The job of transfer is to relate either lexical entries or language-specific predicates in source and target languages. This process may be aided by the use of a domain model which encodes real world knowledge (at least for a restricted domain system such as Verbmobil) but this is assumed to be language-independent. We assume that the transfer component directly relates source and target representations rather than going via the domain model, and thus the domain model is an aid to translation, rather than an inherent part. There are two main reasons for this:

1. It is implausible that any domain model for Verbmobil would be able to cover all the situations that arise in the dialogues, especially those which are reasons or excuses for disliking a particular time for a meeting: appointments at the dentist, sports and so on. These situations are outside the domain proper. However, the system has to be able to translate them, though a lower reliability is likely to be acceptable. For crude translation at least, all that is needed is an equivalence relationship between the lexical entries or lexical semantic structures.
2. The modularity of the system is improved if the domain model and the transfer module are distinct. Changing domains should ideally involve adding to the transfer rules and

refining their conditions, rather than completely rewriting them, as would be necessary if the domain model was central (see the earlier discussion of knowledge-based MT).

**Implementation and control** Simplicity and efficiency are both clearly desirable. We will not attempt to give complexity results here, because it is unclear whether they have much relationship to the practical efficiency. Any approach has the fundamental difficulty that the number of results of transfer is potentially exponential: e.g. if each of $n$ units (signs or semantic structures) has two transfer rules which may apply, there are $2^n$ possible results, which is bad news for any realistic value of $n$. However, there is a trade-off between efficiency and rule extensibility if an attempt is made to control this in the transfer rules directly. Exactly the same arguments about declarativity apply as in grammar rule development. We will make some limited remarks on control in §7.3.

**Acquisition** How easy is it to write transfer equivalences?

Writing transfer rules is inherently complex: it requires knowledge of both source and target language, familiarity with the formalism in which the rules are expressed, and some knowledge of the source and target grammars, or at least of their semantic component. It is important to make the acquisition of transfer rules as easy as possible, given these constraints. Obviously this is interrelated with several of the issues listed above.

For the current Verbmobil system, totally manual acquisition of transfer equivalences is most appropriate, since the size of the lexicon is small. It is likely that essentially manual construction will always be required for the core vocabulary, but in the long term, semi-automatic acquisition of transfer rules is desirable for a robust system, because of the need to deal with unforeseen input. Even with the limited vocabulary currently assumed for Verbmobil, it is possible that a word may be used in a sense which has not been considered by the developers of the transfer rules. Techniques for automatic translation acquisition have, in general, either made use of aligned bilingual corpora or machine readable bilingual dictionaries. Since there is no sufficiently extensive corpus for the Verbmobil domain, the only possibility is to make use of machine readable dictionaries, and some brief remarks are made about this in the subsequent sections.

The aim with many recent approaches to transfer has been to achieve declarativity and reversibility for the same reasons that it is now generally accepted that grammatical formalisms should have these properties. A side effect of this is that it should be possible to evaluate transfer formalisms abstractly, without considering details of the implementation, just as it is possible to evaluate grammatical formalisms without considering the particular parsing strategy chosen. (It is, of course, important that an efficient implementation should be possible in principle). Similarly, concerns of modularity and expression of generalisations apply as much to the expressions of translation equivalence as they do to monolingual lexicons.

## 3.2 Some approaches to semantic transfer

In the next sections, we compare in detail three approaches to the statement of translation equivalences in semantically based transfer. These can be taken as representative of three styles, according to the classification we gave in §2: unification-based syntactic/semantic transfer, lexicalist transfer, and ‘pure’ semantic transfer. The specific systems we discuss are the Verbmobil MDS system, the ACQUILEX tlink mechanism (Copestake, 1993) and transfer using minimal recursion semantics (MRS) (Copestake *et al* 1995). The comparison is at a theoretical level of comparing the formalisation of transfer equivalences, rather than comparing systems. None of these approaches have been used in large scale MT so far. In some ways this is advantageous, since it appears that it is possible (with sufficient effort) to make almost any MT approach work: the resulting systems would be expected to differ considerably in ease of update and maintenance and so forth, but this is extremely difficult to evaluate. It does seem to be generally true that MT systems are extremely difficult to scale up, and our aim here in doing a largely theoretical study is to try and establish in advance where such problems are likely to arise.

The MDS system is the only one which can be considered as a full transfer component — it has been used to construct translations of a number of real dialogues collected for Verbmobil and is the only one of the three approaches which has been used to treat an independently constructed corpus, even though on a rather limited scale. The ACQUILEX tlink mechanism was primarily intended as a technique for representing multilingual lexical knowledge bases, rather than as an MT system itself. However, it is quite closely based on Shake-and-Bake, which is currently being used to develop an MT system at Sharp Laboratories. Tlinks themselves have been used by Trujillo (1995) in a small MT system. MRS is an approach to semantic representation designed to facilitate transfer and generation, particularly with respect to the use of underspecification and the problem of logical form equivalence considered earlier. It is currently being used for the semantics of both English and German grammars in the context of Verbmobil, although the German (IBM Heidelberg) version differs in some respects from the one we will consider here. MRS would support forms of transfer other than the one sketched here, which was intended mainly as a demonstration of the utility of MRS. In fact, as we will see, a transfer system might well be devised which would take elements from each approach.

All the approaches allow one to state equivalences between semantic representations in the source and target language systems. In very abstract terms, they all allow statements such as the following:

$$\mathit{schlecht}'(x) \iff \mathit{bad}'(x)$$

All assume an underlying typed constraint-based formalism, and have a lexicalist orientation, though to varying degrees. But they differ in their answers to the following questions:

1. What are the units to which the transfer rules apply?

2. How are the results of transfer combined?

For the MDS the basic units of transfer are lexical signs, which are accessed via a procedure which operates on the parsed structure. Combinations of lexical signs can be described, but there are limitations on how straightforwardly these can be stated. The Verbmobil semantic construction operations are used to combine the results of transfer into a single DRS which can be input to generation.

Shake-and-Bake is based on the transfers of groups of lexical signs. Arbitrary combinations of signs are possible, so the signs are thought of as forming a bag on which transfer operates to give a bag of target signs. No logical form is produced prior to the 'generation' stage. Instead the constraints are provided by the target bag of lexical signs, and generation can be viewed as a process of sorting an unordered lexical input. The original algorithm for generation was factorial in complexity but Poznanski *et al* (1995) describe a polynomial algorithm, which relies on some (apparently quite natural) constraints on the grammar. The ACQUILEX tlink formalism described in Copestake (1993) is a variant of the lexicalist approach which makes it possible to state additional constraints on the syntactic relationship between the signs being transferred and to state generalisations about classes of translation equivalence.

MRS transfer operates on arbitrary sets of semantic relations, with no access to the syntax. Because the representation is flat, accessing individual elements is trivial. The target logical form is constructed by simply appending the results of transfer.

## 4 Transfer in the MDS

In this section we discuss the transfer component which was implemented for the Verbmobil demonstrator. This had to be designed to fit in with other parts of the Verbmobil system, which somewhat restricted the possibilities. The approach taken in the MDS can be characterised as a variant of linked syntactic/semantic transfer. The input to the transfer process is the result of parsing a German sentence and instantiating its semantics, and the transfer process walks over the HPSG parse tree, applying transfer rules to the lexical signs. All the rules are expressed in the unification-based grammar formalism STUF (Dörre and Seiffert, 1991). The following is an example of one of the tree-walking rules:

```
tau(phrase_s &
    dtrs: (head_adj_struc &
           adj_dtr: #SemHead),#Args) => tau(#SemHead,#Args).
```

Such rules apply to the different types of syntactic structure in the parse tree (a head-adjunct structure in this case), and recurse on the subcomponents. Eventually the process bottoms out at lexical signs:

```
tau(word_s & #Sign,#Args) => tau_lex(#Sign,#Args).

tau_lex(pred_name(#SourcePred) & #Sign,#Args) =>
        tau_lex(#SourcePred,#TargetPred,#SemArgs,#Args,#Sign) &
        sem_lex(#TargetPred,#SemArgs).
```

The effect of this general rule is to apply a specific transfer rule which will instantiate the value of `#TargetPred` and `#SemArgs`, which will enable `sem_lex` to build the semantic structure for the transferred predicate.

A particular instance of a rule is given below:

```
tau_lex(schlecht,bad,[tau(#NP,#Args)],#Args,ad_syn(#NP)) => sem_t.
```

This rule allows *schlecht* to be translated as *bad* by giving the correspondence between the German and the English predicates and their respective arguments. The recursion `tau(#NP,#Args)` instantiates the arguments for `sem_lex`. The fourth argument to `tau_lex`, `#Args`, will be discussed below. The sign as a whole has to unify with `ad_syn(#NP)` so this has the effect of stipulating a syntactic condition on the translation — that *schlecht* is being used as an adjective.

To give an idea of how this rule is used in translation, consider the simple example in (14a):

(14) a Das ist schlecht.

b That is bad.

The German semantic structure for (14a) is as shown in Figure 3. This structure is a DRS, but we have used the low-level representation here because this makes the operation of the transfer rules clearer. Of course, the transfer process sees more information than this since it works on the full result of the parse, and thus has access to the syntactic structure as well as to the semantics. We do not show the full parse structure here, however, because it is complex and the details are not

```
[(sem:(sem_t &
       lambda:[] &
       drs:(drs_t &
            dom:[(A &
                  marker &
                  sort:statisch_c)] &
            conds:[(B &
                    supp_condition &
                    supp_drs:(drs_t &
                              dom:[] &
                              conds:[(alfa_condition &
                                      alfa_arg:(C &
                                                marker &
                                                snumber:individual) &
                                      alfa_type:demonstr &
                                      alfa_restr:(drs_t &
                                                  dom:[C] &
                                                  conds:[])),
                                     (dimen_condition &
                                      dimen_inst:C &
                                      dimen_pred:schlecht &
                                      dimension:quality)]) &
                    supp_state:A)]) &
       quants:[] &
       ip:(ip_t &
           cond:B &
           idx:A) &
       persp:(persp_t &
              p_inst:A))]
```

Figure 3: Semantics for *Das ist schlecht* omitting tense information

important, since the transfer rules use high level predicates such as `trans_syn`. Thus the transfer rule writer should not need to know the details of the syntactic representation, and a reasonable degree of modularity is maintained.

For simplicity we have also omitted the representation of tense: the approach to tense is described in detail in Butt (1995). In addition to the rule for *schlecht* the transfer rules shown below are used in the translation of (14):

```
tau_lex(sein1,be1,[tau_tense(#Tense),tau(#O)],[],
        trans_syn(#S,#O & mod([#S]),#Tense)) => sem_t.

tau_lex(pronoun,pronoun,[#Marker,#AlfaType],[],
        sem:cond_sem(alfa_cond(#Marker,#AlfaType,[]))) => sem_t.
```

The first rule is for predicative uses of *sein*. Note the translation of the tense stipulated by `tau_tense(#Tense)`. Again, a syntactic condition is stipulated on the rule by the final argu-

```
*c_id:c1027
 pred:alfa
 conc:_6375477
 m_id:m561
 args:[mod_rel:[]]
 mods:[type_info(alfa,demonstr)]

 *c_id:c1028
 pred:bad
 conc:_6376517
 m_id:m561
 args:[dimension: (mood_dim;quality)]
 mods:[]

*m_id:m561
 sort:_6375311
 numb:individual
 gend:_6375313
 conc:_6375314
 c_id:_6375315
 mods:[]
 args:[c1028,c1027]
```

Figure 4: Intermediate results of transfer (omitting tense)

ment (which also instantiates `#S` and `#O`). The second rule transfers pronouns: these are, in effect, analysed using an interlingua, as the information about pronoun type and so on is carried over directly from German to English.

During actual translation, these rules will apply to an instantiated structure, in such a way that the coindexation between the arguments of the target language predicates is preserved on transfer to the source language predicates. For example, the semantics for the English predicate `bad` are specified as follows:

```
sem_lex(bad,#Args) => ad_dimen_sem1(bad,quality;mood_dim,#Args).
```

The intermediate results after the application of the transfer rules for (14a) are shown in Figure 4. These results are "stitched together" by the semantic composition process. The English semantic structure that results is shown in Figure 5.

Signs corresponding to arguments can straightforwardly be accessed by transfer rules operating on a verb etc, making it easy to state rules which transfer source verb-argument combinations. However, when it is necessary to apply a rule to a verb-adjunct combination, the treatment becomes more complex. Consider the example in (15a).

(15) a Das paßt schlecht bei mir

b That does not suit me at all

```
[(sem:(sem_t &
          lambda:[] &
          drs:(drs_t &
               dom:[(A &
                     marker &
                     sort:statisch_c)] &
               conds:[(B &
                       supp_condition &
                       supp_drs:(drs_t &
                                 dom:[] &
                                 conds:[(alfa_condition &
                                         alfa_arg:(C &
                                                   marker &
                                                   snumber:individual) &
                                         alfa_type:demonstr &
                                         alfa_restr:(drs_t &
                                                     dom:[C] &
                                                     conds:[])),
                                        (dimen_condition &
                                         dimen_inst:C &
                                         dimen_pred:bad &
                                         dimension:(mood_dim ; quality))]) &
                       supp_state:A)]) &
          quants:[] &
          ip:(ip_t &
              cond:B &
              idx:A) &
          persp:(persp_t &
                 p_inst:A))]
```

Figure 5: Semantics for generating *That is bad* omitting tense information

Here *passen* (*suit*) is analysed as an intransitive, with the experiencer being realised as an adjunct related to the verb by *bei*. As a further complication, *schlecht* cannot be translated as *badly* here.[6] The transfer component thus has to make *not at all* or *not well* available as alternatives. The rules used are shown below:

```
tau_lex(passen,suit,[tau_tense(#Tense), tau(#S), #OSem],[#OSem],
        intrans_syn(#S,#Tense)) => sem_t.

tau_lex(bei1,dummy_pred,[tau(#VP,[tau(#NP)|#Args])],#Args,
        prep_syn(#VP&sem:cond_sem(basic_cond(gehen2;
passen;kommen_in_frage,marker,list)),#NP)) => sem_t.

tau_lex(schlecht,negation,[focus_adv_sem1(at_all, grad, [tau(#VP,#Args)])],
[at_all|#Args],adv_syn(#VP&lex_pred(passen;passen1;gehen3;gehen2)))
=> sem_t.
```

Because *bei mir* is analysed as an adjunct, the translation of the NP is added to the list `#Args`, so that it will be available when the verb itself is translated. Thus the fourth argument to the `tau_lex` rule for *passen* is instantiated to a list containing the translation of *mir*, ensuring that the correct coindexation will take place, so that this will form the object argument of the transitive verb *suit*. No translation of *bei* itself is produced by this rule. The rule is restricted so that it only applies in a limited range of cases: e.g. when the verb is *gehen* (in the sense which can be translated as *be acceptable*) or *passen*. A rather similar treatment is used for *schlecht*, which adds `at_all` to the value of `Args`.

Finally, we should mention the interaction with the domain model. This is illustrated below, where three alternative translations of *Termin* are given:

```
tau_lex(termin,appointment,[],[],
  pred_sort(subsumes(veranstaltung_c))& #SourceSign) => tau_nn_nn(#SourceSign).
tau_lex(termin,date,[],[],
  pred_sort(subsumes(zeit_punkt_c))& #SourceSign) => tau_nn_nn(#SourceSign).
tau_lex(termin,appointment_slot,[],[],
  pred_sort(subsumes(zeit_periode_c))& #SourceSign) => tau_nn_nn(#SourceSign).
```

The three usages are disambiguated by reference to the domain model concepts: `veranstaltung_c` (arrangement), `zeit_punkt_c` (time point), `zeit_periode_c` (time period).

Considering the criteria we gave in the previous section, we can make the following observations:

**Coverage** This may not be restricted in principle, but the solution that had to be adopted to the example given above where *passen bei mir* was translated as *suit me*, is somewhat messy. This sort of transfer rule could be needed quite frequently.

**Modularity** Even though syntactic information is accessed, this is achieved in a reasonably abstract way, so the transfer rule writer does not need detailed knowledge of the source and target grammars. The use of the semantic construction procedures means that there is no need to directly tune the output of transfer to the target grammar.

**Independence** Transfer rules have to interact in quite complex ways in order to ensure that the correct values are passed around.

[6] As we will mention below, this is actually due to a monolingual restriction, since *suit* is a positive psychological predicate in English which does not normally take a negative attitude denoting adverbial such as *badly*.

**Underspecification** The approach would work with an underspecified semantic representation of the type discussed in §2.2, but as described it is operating on more fully specified representations.

**Generalisation** The formalism allows generalisations to be made over classes of transfer rules, but as the rules are currently expressed there is some redundancy.

**Interaction with generation** The output of transfer is a logical form, which could be underspecified to allow lexical choice in generation, or even be reformulated by a strategic generator.

**Interaction with domain model etc** As illustrated, the system can make the choice of transfer rule dependent on conditions which are checked with respect to the domain model.

**Implementation and control** The tree-walking technique is efficient, but this is achieved at the expense of flexible statement of some transfer relationships. It is possible to state a variety of conditions which have to be met for a rule to be applicable, which is very useful. However there is no clear distinction between such conditions and parts of the rule which have to be evaluated in order to instantiate some variable. Because the rules are written in STUF, evaluation order can be controlled: it is therefore possible to have rules which act as defaults.

**Acquisition** Many of the transfer rules are quite straightforward, but others could only be written by someone with a quite detailed knowledge of the STUF formalism.

# 5 Shake-and-Bake and tlinks

In this section, we discuss lexicalist approaches to transfer in some detail. To start with we give an introduction to Shake-and-Bake (Whitelock, e.g. 1992) which was the first approach in this category and then describe the ACQUILEX translation link (tlink) formalism, which was actually designed for use in lexical knowledge bases (Sanfilippo *et al* (1992), Copestake (1993), Copestake *et al* (1995)).

## 5.1 Shake-and-Bake

As originally described, Shake-and-Bake depends on relating monolingual lexical signs described within a lexicalist grammatical framework. The only information which is actually transferred is the values of the indices which become instantiated during parsing (though see below). The 'generator' is given a bag of lexical signs with their semantic arguments instantiated and actually generates by parsing, trying all possible orderings, accepting only those with the appropriate coindexation. The advantage of this approach is that the transfer component contains no information about the monolingual grammar, since it merely relates existing lexical entries, and that the problem of LF equivalence is to a large extent circumvented, because the transfer component only indicates the relation of the lexical signs by coindexation. One major disadvantage of Shake-and-Bake as originally described is lack of efficiency, since the generator/parser has to consider a number of possibilities which is factorial in the number of signs in the target sentence. However, Poznanski *et al* (1995) show that polynomial complexity in Shake-and-Bake generation can be achieved with some restrictions on the grammar.

Consider the trivial example we used in the previous section:

(16) a Das ist schlecht

b That is bad

$$
\left[\begin{array}{ll}
\textsc{phon} & \left\langle \textit{das ist schlecht} \right\rangle \\
\textsc{sem} & \boxed{1} \left[\begin{array}{ll}
\textsc{pred} & \textit{and} \\
\textsc{arg1} & \left[\begin{array}{ll} \textsc{pred} & \textit{deixis} \\ \textsc{arg} & \boxed{2} \end{array}\right] \\
\textsc{arg2} & \boxed{4} \left[\begin{array}{ll}
\textsc{pred} & \textit{and} \\
\textsc{arg1} & \left[\begin{array}{ll} \textsc{pred} & \textit{schlecht\_state} \\ \textsc{event} & \boxed{3} \\ \textsc{arg1} & \boxed{2} \end{array}\right] \\
\textsc{arg2} & \boxed{5} \left[\begin{array}{ll} \textsc{pred} & \textit{present} \\ \textsc{arg1} & \boxed{3} \end{array}\right]
\end{array}\right]
\end{array}\right] \\
\textsc{dtrs} & \left[\begin{array}{ll}
\ldots & \left[\begin{array}{ll} \textsc{phon} & \left\langle \textit{das} \right\rangle \\ \textsc{sem} & \boxed{1} \end{array}\right] \\
\ldots & \left[\begin{array}{ll} \textsc{phon} & \left\langle \textit{schlecht} \right\rangle \\ \textsc{sem} & \boxed{4} \end{array}\right] \\
\ldots & \left[\begin{array}{ll} \textsc{phon} & \left\langle \textit{ist} \right\rangle \\ \textsc{sem} & \boxed{5} \end{array}\right]
\end{array}\right]
\end{array}\right]
$$

Figure 6: Schematic structure for *Das ist schlecht*

A sketch of the instantiated structure for (17a) is given in Figure 6. We have assumed a different treatment of predicative adjectives from that in the MDS, in order to make the treatment clearer. The lexical signs, now with instantiated semantics, are treated as forming a 'bag', to which the lexical equivalences shown below can be applied:

$$
\left[\begin{array}{ll}
\textsc{phon} & \left\langle \textit{das} \right\rangle \\
\textsc{sem} & \left[\begin{array}{ll}
\textsc{pred} & \textit{and} \\
\textsc{arg1} & \left[\begin{array}{ll} \textsc{pred} & \textit{deixis} \\ \textsc{arg} & \boxed{2} \end{array}\right] \\
\textsc{arg2} &
\end{array}\right]
\end{array}\right]
\Longleftrightarrow
\left[\begin{array}{ll}
\textsc{phon} & \left\langle \textit{that} \right\rangle \\
\textsc{sem} & \left[\begin{array}{ll}
\textsc{pred} & \textit{and} \\
\textsc{arg1} & \left[\begin{array}{ll} \textsc{pred} & \textit{deixis} \\ \textsc{arg} & \boxed{2} \end{array}\right] \\
\textsc{arg2} &
\end{array}\right]
\end{array}\right]
$$

$$
\left[\begin{array}{ll}
\textsc{phon} & \left\langle \textit{schlecht} \right\rangle \\
\textsc{sem} & \left[\begin{array}{ll}
\textsc{pred} & \textit{and} \\
\textsc{arg1} & \left[\begin{array}{ll} \textsc{pred} & \textit{schlecht\_state} \\ \textsc{event} & \boxed{3} \\ \textsc{arg1} & \boxed{2} \end{array}\right] \\
\textsc{arg2} &
\end{array}\right]
\end{array}\right]
\Longleftrightarrow
\left[\begin{array}{ll}
\textsc{phon} & \left\langle \textit{bad} \right\rangle \\
\textsc{sem} & \left[\begin{array}{ll}
\textsc{pred} & \textit{and} \\
\textsc{arg1} & \left[\begin{array}{ll} \textsc{pred} & \textit{bad\_state} \\ \textsc{event} & \boxed{3} \\ \textsc{arg1} & \boxed{2} \end{array}\right] \\
\textsc{arg2} &
\end{array}\right]
\end{array}\right]
$$

$$
\left[\begin{array}{ll}
\textsc{phon} & \left\langle \textit{ist} \right\rangle \\
\textsc{sem} & \left[\begin{array}{ll} \textsc{pred} & \textit{present} \\ \textsc{arg1} & \boxed{3} \end{array}\right]
\end{array}\right]
\Longleftrightarrow
\left[\begin{array}{ll}
\textsc{phon} & \left\langle \textit{is} \right\rangle \\
\textsc{sem} & \left[\begin{array}{ll} \textsc{pred} & \textit{present} \\ \textsc{arg1} & \boxed{3} \end{array}\right]
\end{array}\right]
$$

The resulting bag of target lexical entries is shown below:

$$\left\{ \begin{bmatrix} \textsc{phon} & \langle \textit{is} \rangle \\ \textsc{sem} & \begin{bmatrix} \textsc{pred} & \textit{present} \\ \textsc{arg1} & \boxed{3} \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \textsc{phon} & \langle \textit{that} \rangle \\ \textsc{sem} & \begin{bmatrix} \textsc{pred} & \textit{and} \\ \textsc{arg1} & \begin{bmatrix} \textsc{pred} & \textit{deixis} \\ \textsc{arg} & \boxed{2} \end{bmatrix} \\ \textsc{arg2} & \end{bmatrix} \end{bmatrix}, \begin{bmatrix} \textsc{phon} & \langle \textit{bad} \rangle \\ \textsc{sem} & \begin{bmatrix} \textsc{pred} & \textit{and} \\ \textsc{arg1} & \begin{bmatrix} \textsc{pred} & \textit{bad\_state} \\ \textsc{event} & \boxed{3} \\ \textsc{arg1} & \boxed{2} \end{bmatrix} \\ \textsc{arg2} & \end{bmatrix} \end{bmatrix} \right\}$$

These are then combined by the target grammar, a process which can be regarded as a form of constraint resolution or of parsing with unordered lexical items, rather than conventional generation from logical form. The result is the structure shown in Figure 7: a logical form for the target sentence is produced, but this is a side effect of producing the output string.

$$\begin{bmatrix} \textsc{phon} & \langle \textit{that is bad} \rangle \\ \textsc{sem} & \boxed{6} \begin{bmatrix} \textsc{pred} & \textit{and} \\ \textsc{arg1} & \begin{bmatrix} \textsc{pred} & \textit{deixis} \\ \textsc{arg} & \boxed{2} \end{bmatrix} \\ \textsc{arg2} & \boxed{7} \begin{bmatrix} \textsc{pred} & \textit{and} \\ \textsc{arg1} & \begin{bmatrix} \textsc{pred} & \textit{bad\_state} \\ \textsc{event} & \boxed{3} \\ \textsc{arg1} & \boxed{2} \end{bmatrix} \\ \textsc{arg2} & \boxed{8} \begin{bmatrix} \textsc{pred} & \textit{present} \\ \textsc{arg1} & \boxed{3} \end{bmatrix} \end{bmatrix} \end{bmatrix} \\ \textsc{dtrs} & \begin{bmatrix} \ldots & \begin{bmatrix} \textsc{phon} & \langle \textit{that} \rangle \\ \textsc{sem} & \boxed{6} \end{bmatrix} \\ \ldots & \begin{bmatrix} \textsc{phon} & \langle \textit{bad} \rangle \\ \textsc{sem} & \boxed{7} \end{bmatrix} \\ \ldots & \begin{bmatrix} \textsc{phon} & \langle \textit{is} \rangle \\ \textsc{sem} & \boxed{8} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Figure 7: Schematic structure for *Das ist schlecht*

Note that, even though we used an interlingual representation for *das* and *that*, it was still necessary to explicitly specify the lexical equivalence, unlike in the MDS. We have glossed over the treatment of morphology but in Shake-and-Bake it is usual to assume an inflectional morpheme which is simply treated as another lexical item (although there are other possible approaches).

More complex cases involve relating sets of lexical items. For example, in order to represent the translation equivalence between *passen bei NP* and *suit NP* described in the previous section, and assuming the same analysis, the following relation could be used:

$$
\left\{
\left[\begin{array}{ll}
\textsc{phon} & \left\langle \textit{passen} \right\rangle \\
\textsc{sem} & \left[\begin{array}{ll}
\textsc{pred} & \textit{and} \\
\textsc{arg1} & \left[\begin{array}{ll}
\textsc{pred} & \textit{passen} \\
\textsc{event} & \boxed{3} \\
\textsc{arg1} & \boxed{2}
\end{array}\right] \\
\textsc{arg2} &
\end{array}\right]
\end{array}\right],
\left[\begin{array}{ll}
\textsc{phon} & \left\langle \textit{bei} \right\rangle \\
\textsc{sem} & \left[\begin{array}{ll}
\textsc{pred} & \textit{and} \\
\textsc{arg1} & \left[\begin{array}{ll}
\textsc{pred} & \textit{bei} \\
\textsc{arg1} & \boxed{3} \\
\textsc{arg2} & \boxed{1}
\end{array}\right] \\
\textsc{arg2} &
\end{array}\right]
\end{array}\right]
\right\}
\Longleftrightarrow
$$

$$
\left[\begin{array}{ll}
\textsc{phon} & \left\langle \textit{suit} \right\rangle \\
\textsc{sem} & \left[\begin{array}{ll}
\textsc{pred} & \textit{and} \\
\textsc{arg1} & \left[\begin{array}{ll}
\textsc{pred} & \textit{suit} \\
\textsc{event} & \boxed{3} \\
\textsc{arg1} & \boxed{2} \\
\textsc{arg2} & \boxed{1}
\end{array}\right] \\
\textsc{arg2} &
\end{array}\right]
\end{array}\right]
$$

The only relationship between the signs for *passen* and *bei* is that indicated by coindexation of the event variable.

Now consider the more complex example which we discussed in the previous section:

(17) a Das paßt schlecht bei mir

b That does not suit me well

We have changed the translation slightly for ease of exposition. The Shake-and-Bake style of representing equivalences runs into some problems in relating *schlecht* to *not well*, since *not* and *well* cannot be related by coindexation in a conventional semantic representation. In the MDS, it was possible to make use of a complex equivalence which allowed for dependencies between parts of the verb phrase which were only indirectly related. The MDS expression did make reference to the verb *passen*, but this was not essential, and in fact, as we will discuss in §7.3 it would be more appropriate not to limit this transfer rule to particular verbs, since the unacceptability of the literal translation *that suits me badly* is due to monolingual considerations. To achieve the same correspondence in Shake-and-Bake however, the inclusion of the verb in the source bag is obligatory, because it is only via the verb that *not* and *well* can be related. This leads to an unnecessarily complex and specific transfer rule where the source bag has to contain the lexical entries for *passen*, *bei* and *schlecht* and the target bag contains *suit*, *not* and *well*.

In fact there is a general problem with Shake-and-Bake style mechanisms which is that they need additional machinery to deal with any form of higher-order semantics. For example, suppose the semantics of (18a) corresponds to the linearised representation shown in (18b):

(18) a very aggressive dog

b $[x][\text{very}(\text{aggressive})(x)]$

In either case, the corresponding FS does not have a suitable index with which to relate *very* and *sehr* because their arguments are language-specific predicates.

Note that we cannot simply create a flat structure where *very* takes the index $x$ as an argument, because this would prevent differentiation between (19a) and (19b) since they would both have the representation shown in (19c).

(19) a small very aggressive dog

b aggressive very small dog

c $[x][\text{small}(x) \wedge \text{very}(x) \wedge \text{aggressive}(x)]$

Since DRT is a first order theory, this particular problem is not going to arise in Verbmobil, but the difficulty of representing the scope of negation is a problem, as in the *schlecht≈not good* example. There are various ways round this, one of which would be to give the predicate itself an index which could be shared between SL and TL representations: this would be analogous to the MRS handel mechanism discussed in §6.

### 5.2 Tlinks

The ACQUILEX translation link (tlink) language is a way of describing lexical equivalences, which was intended to be utilisable by a variety of approaches to MT. Since the aim was to build a multilingual lexical knowledge base (LKB) it was necessary to come up with a formalism that directly related lexical entries, which resulted in a technique for representing lexical equivalences which had close affinities with that used in Shake-and-Bake. Rather than concentrating on developing an efficient transfer mechanism, the aim was to make the LKB representation sufficiently flexible to be able to represent generalisations about translation equivalences. This involved a combination of techniques:

1. Translation equivalences (tlinks) are themselves typed feature structures, so generalisations can be made over classes of equivalence, making it unnecessary to keep repeating the feature paths, for example. This makes it easier to update the bilingual lexicon, if the structure of a monolingual lexicon is changed. For example, the *schlecht≈bad* relationship could be encoded as follows:

   ```
   schlecht_prd / bad_prd:
   pred-adj-tlink.
   ```

   Here `schlecht_prd` and `bad_prd` refer to the entire lexical entries, which then can instantiate the feature structure given by *pred-adj-tlink*. Even if the semantic representation of predicative adjective were changed, all that is necessary is to update the type *pred-adj-tlink* for this description to continue to apply.

2. A level of indirection was used so that the translation equivalences could always be related to lexical entries. This is shown diagrammatically for the *schlecht≈bad* tlink in Figure 8, although for a simple tlink, such as this one, the relationship between the lexical entries and the forms to be linked is simply identity. The indirection is significant for more complex relationships. Consider a link between a singular and a plural noun (e.g. *furniture≈Möbel*): assuming that lexical entries are stored in the singular form, the relationship between the lexical entry and the target sign is simply the rule for plural formation. This feature is also useful for phrasal tlinks such as that for *passen bei≈suit*. The rule could simply be stated as follows:

   ```
   passen_1 bei_1 / suit_1 :
   intrans-verb-adjunct-to-trans-tlink.
   ```

   The structure that this would expand to is shown diagrammatically in Figure 9.

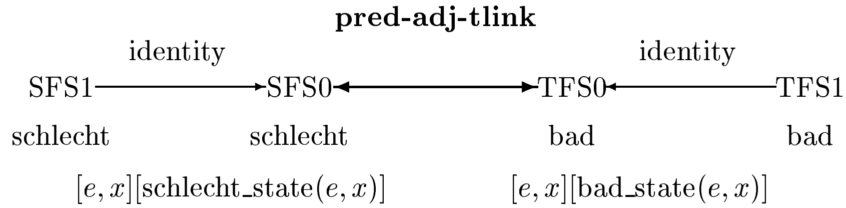


Figure 8: Figurative representation of translation link relating *schlecht* and *bad*

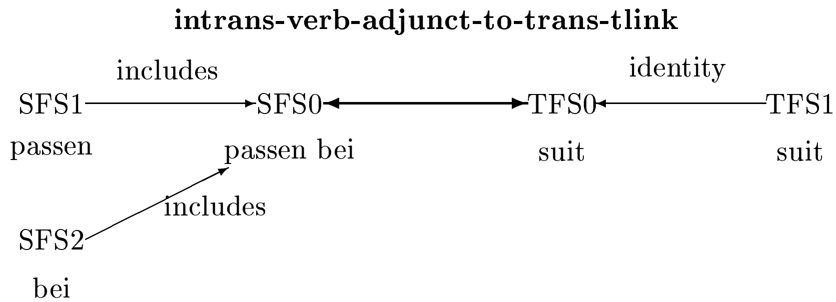


Figure 9: Tlink relating *passen bei* and *suit*

3. Although it is possible to represent the phrasal relationships as simply involving sets of lexical signs, related only by coindexation, is is also possible to express extra constraints on how the elements in the bag can be combined. However, doing this with full generality (while retaining the advantages of the Shake-and-Bake mechanism with respect to the logical form equivalence problem) comes at the cost of introducing an operation equivalent to LFG functional uncertainty (indicated by `includes` in Figure 9) and the author no longer considers this a particularly promising approach.

4. The tlink mechanism allows rules as well as lexical entries to be related cross-linguistically, using *tlink-rules*. For example, causatives of movement verbs such as *fly* in English (*fly the kite*) can be related to an explicit causative construction in the Romance languages (e.g. *fece volare l'acquilone*), see Copestake and Sanfilippo (1993). This mechanism can also be used when relating processes which are more directly equivalent in the two languages. As a simple example, assume that in both the German and the English grammars, predicative adjectives are related to attributive ones by lexical rule. A tlink rule relating these lexical rules would have the result that the existence of a tlink between attributive adjectives would automatically cause the predicative forms of those adjectives to be related as well, as indicated diagrammatically in Figure 10. Another case where tlink rules could be used is the metonymic interpretation of *Stuttgart* as *meeting in Stuttgart*, mentioned in §2.2. Tlink rules avoid the necessity to directly specify very large numbers of predictable translation equivalences such as *fly (causative)*≈*fare volare*, *run (causative)*≈*fare correre*, *march (causative)*≈*fare marciare* etc. However, to use the Shake-and-Bake algorithm, all these equivalences have to be compiled out by the system, so there is no gain in efficiency of implementation.

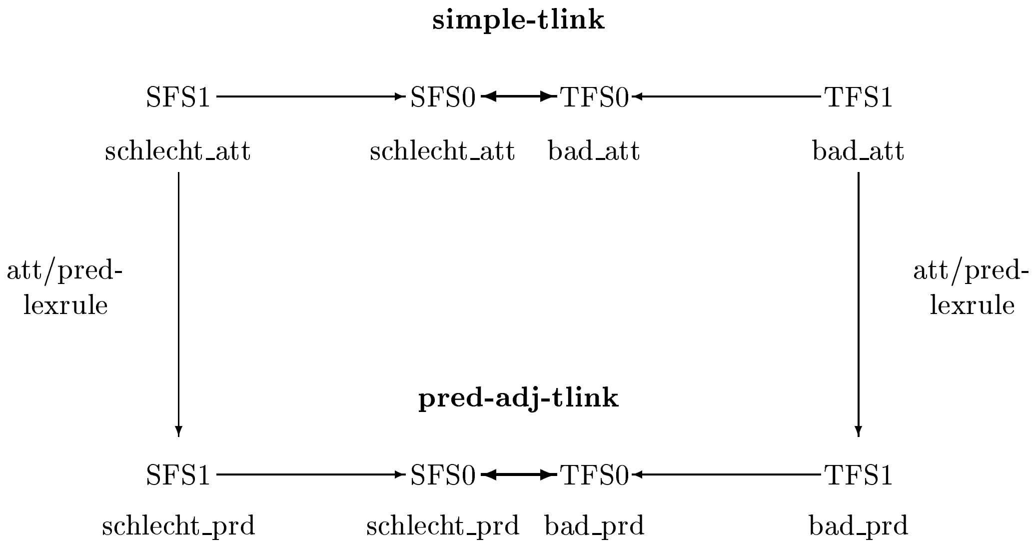


Figure 10: Tlink rule relating attributive and predicative adjectives, shown instantiated with a tlink for *schlecht≈bad*

### 5.3 Summary

For further details of Shake-and-Bake and the tlink mechanism, see the papers cited. Here we summarise the advantages and disadvantages, according to the criteria we discussed earlier:

**Coverage** Some phenomena cannot be covered adequately without extensions to the formalism. Shake-and-Bake requires a lexicalist approach, since if semantic information is introduced by grammar rules, it cannot be related by the lexical equivalences. This considerably restricts the class of grammars which can be used, and even precludes using some versions of HPSG which utilise the phrasal sort hierarchy to represent constructions (e.g. Sag, 1995). Phrasal tlinks could be used, but the Poznanski *et al* generation algorithm would not then be applicable and generation would be intractable.

**Modularity** Although syntactic conditions on transfer rules could be stated, the formalism simply requires knowledge of the semantics. The target grammar can be developed independently from the set of transfer rules because the logical form equivalence problem does not arise.

**Independence** If rules are fully instantiated, they are mutually independent. However, the number of fully instantiated rules can become very large, and if techniques such as tlink rules are used, interdependence between rules has to be considered.

**Underspecification** The Shake-and-Bake formalism assumes that the output of transfer is a bag of lexical entries. It is possible to relax this, so that the output contains some underspecified lexical entries, but the efficiency will be affected.

**Generalisation** A number of techniques for generalising over classes of transfer rules have been discussed, especially in the ACQUILEX tlink papers.

**Interaction with generation** Standard approaches to generation from a logical form cannot be used. There is no straightforward way of integrating a strategic generation component.

**Interaction with domain model etc** Interaction with a domain model which operated by specialising sorts in the typed feature structure would be straightforward. For example, analogous rules for those given in the MDS for the translation of *Termin* could be written in Shake-and-Bake. But there are difficulties in implementing more complex schemes, which are discussed in §7, below.

**Implementation** Poznanski *et al* (1995) describe a polynomial time Shake-and-Bake generator. Most of the papers on Shake-and-Bake and related systems do not mention control of application of transfer rules, but some control scheme is clearly required.

**Acquisition** Coding translation equivalences in Shake-and-Bake and in the ACQUILEX tlink mechanism is simpler than in the MDS. Automatic acquisition of simple tlinks has been demonstrated (Copestake *et al*, 1995) but the techniques described there could be used with other approaches to transfer representation, especially those that support generalisations of the type covered by tlink-rules.

# 6 Transfer using Minimal Recursion Semantics (MRS)

The description of MRS which follows is somewhat longer than the previous two sections, because we are describing both an approach to semantic representation within a unification-based formalism and how it can be utilised for transfer. Much of the following discussion is taken from Copestake *et al* (1995). Briefly, minimal recursion semantics (MRS) is a framework for semantic representation within HPSG, which can be thought of as a meta-level language for describing semantic structures. It supports underspecification, so an MRS description will correspond to a set of object-language expressions. For simplicity, in the examples in this paper we will take the object language to be predicate calculus, but MRS is intended to be compatible with DRT. We discussed the advantages of allowing underspecification in a semantic representation for transfer in §2.2. The other novel feature of MRS is that it is a flat semantic representation. In §2.4 we discussed the problems that arise when driving transfer from a conventional semantic representation: MRS was developed in order to try and alleviate these problems by devising a representation with no recursive structure. Taking the example we discussed earlier, of *novillo≈young bull*, we mentioned that a simplistic encoding of this translation equivalence would not allow for examples such as *young black bull* if this had the following LF:

$$\lambda x[\mathrm{young}(x) \wedge (\mathrm{black}(x) \wedge \mathrm{bull}(x))]$$

In contrast, a flat semantic representation might be:

$$\lambda x \bigwedge[\mathrm{young}(x), \mathrm{black}(x), \mathrm{bull}(x)]$$

A transfer rule can now be written without having to take account of the possible bracketings. For further discussion, see Trujillo (1995). The above example is trivial, of course, but in §6.2, we will show how flat semantics makes it easier to write some more complex transfer rules.

## 6.1 Introduction to MRS

Assuming that it is desirable to reduce the use of recursive structures, it is necessary to find an alternative method of representing scope. For example, ignoring tense for the moment, consider the representation of:

(20) Every tall man is old

In this case there is only one possible scope for *every* which is shown in (21) using generalized quantifiers:

(21) $every(x, man(x) \wedge tall(x), old(x))$

We should, therefore, be able to retrieve this reading unambiguously from the semantic representation that the grammar constructs for this sentence. However, if we have the totally flat structure shown in (22) it is impossible to retrieve the correct reading unambiguously, because we would get the same structure for (23).

(22) $\bigwedge\{every(x), man(x), old(x), tall(x)\}$

(23) Every old man is tall

A flat representation is required which preserves sufficient information about the scope of a quantifier to be able to construct all and only the possible readings for the sentence. Similar remarks apply with respect to the representation of the scope of *not*, *or* and so on.

This effect can be achieved while retaining the flatness of the representation by adding extra variables, which have the effect of capturing scope information. For example, we could represent (23) as (24):

(24) $\bigwedge\{\mathrm{every}(x,1,2), \mathrm{man}_1(x), \mathrm{old}_1(x), \mathrm{tall}_2(x)\}$

These extra variables can be thought of as handles which enable us to 'grab' particular propositions in the flat list (cf the use of *labels* in Frank and Reyle, 1994). In (24) the scoped representation can be reconstructed by replacing the handles in the restriction and body arguments of *every* with the propositions tagged by those handles. From now on we will drop the use of $\wedge$ and assume implicit conjunction.

In the example above, everything which would have been inside a set of braces in a conventional formula was tagged with the same subscript. Nested quantifiers do not require multiple indices on a single conjunct, since we can trace the nesting via the restriction of the embedded quantifier. The reason why this approach is not just a notational variant of a standard representation is that the handles can be underspecified to represent multiple scopes. For example, the underspecified representation of (25a) would be (25b):[7]

(25) a Every dog chased some cat.

b $\mathrm{every}_1(x,3,n), \mathrm{dog}_3(x), \mathrm{cat}_7(y), \mathrm{some}_5(y,7,m), \mathrm{chase}_4(e,x,y)$

Here $n$ and $m$ stand for variables over handles. The scoped representations would be:

(26) a $\mathrm{every}_1(x,3,4), \mathrm{dog}_3(x), \mathrm{cat}_7(y), \mathrm{some}_5(y,7,1), \mathrm{chase}_4(e,x,y)$
(wide scope *some*)

b $\mathrm{every}_1(x,3,5), \mathrm{dog}_3(x), \mathrm{cat}_7(y), \mathrm{some}_5(y,7,4), \mathrm{chase}_4(e,x,y)$
(wide scope *every*)

[7]We are oversimplifying somewhat here, since in order to be able to represent any information about the relative scope of more than two quantifiers, it is necessary to represent the relative scope of pairs of quantifiers (Frank and Reyle, 1994). This can be accommodated in the MRS representation, but we do not give the details since there are not relevant here.

$$\begin{bmatrix} \text{HANDEL} & \boxed{1} \vee \boxed{5} \\ \text{INDEX} & \boxed{8} \\ \text{LISZT} & \left\langle {}_{\textit{every\_rel}}\begin{bmatrix} \text{HANDEL} & \boxed{1} \\ \text{BV} & \boxed{2} \\ \text{RESTRICTION} & \boxed{3} \end{bmatrix}, {}_{\textit{dog\_rel}}\begin{bmatrix} \text{HANDEL} & \boxed{3} \\ \text{INST} & \boxed{2} \end{bmatrix}, {}_{\textit{some\_rel}}\begin{bmatrix} \text{HANDEL} & \boxed{5} \\ \text{BV} & \boxed{6} \\ \text{RESTRICTION} & \boxed{7} \end{bmatrix}, \right. \\ & \left. {}_{\textit{cat\_rel}}\begin{bmatrix} \text{HANDEL} & \boxed{7} \\ \text{INST} & \boxed{6} \end{bmatrix}, {}_{\textit{chase\_rel}}\begin{bmatrix} \text{HANDEL} & \boxed{4} \\ \text{EVENT} & \boxed{8} \\ \text{ACT} & \boxed{2} \\ \text{UND} & \boxed{6} \end{bmatrix} \right\rangle \end{bmatrix}$$

Figure 11: Unscoped representation for *every dog chased some cat*

MRS proper is defined in terms of feature structures (FSs), rather than the linearized representation shown above. The semantic representation has two parts, CONTENT and CONTEXT, as usual in HPSG, but here we are mainly concerned with the CONTENT value. An MRS expression consists of a structure of type *mrs-struct*, with appropriate features HANDEL and LISZT, which take values of type *handle* and *list* respectively. The feature name LISZT is used to distinguish the non-recursive semantic structure from ordinary lists: the values of LISZTs have to be treated like sets (or, possibly, bags) in some respects, as we will see below. HANDEL is used for the other main feature of the compositional semantics. We adopt the normal convention of writing feature names in (small) capitals and types/sorts in italics. In what follows, we will often use liszt and handel in the normal lower case font to refer to the values of those features.

The sort *mrs-struct* also has the appropriate feature INDEX which plays much the same role as a lambda variable in conventional representations. The value of LISZT is defined to be a flat list of *rel*s (relations), which all have HANDELs and other features depending on their sort. As in Pollard and Sag (1994:chap. 8, sec. 5), the actual relation is indicated by the sort of the rel. Determiners, such as *every*, have rels with appropriate features BV (bound variable) which takes a value of sort *ref-ind*, and RESTR (restriction) which takes a *handle*. MRS assumes a neo-Davidsonian style of representation with explicit event variables. Verb rels have a feature EVENT which takes an event variable. They also have features such as ACT (actor) and UND (undergoer) following Davis (forthcoming). Common nouns have rels with the feature INST which takes a *ref-ind*. An example of the (unscoped) MRS representation for (25a) is shown in Figure 11. Here the variable sorts and the internal structure of the indices are not shown, but only the coindexation between them, indicated as usual by boxed integers. The handel shown at the outer level allows the sentence to be embedded, as in *Sandy said that every dog chased some cat*, for example. Here it is a disjunct of the handels of the quantifiers, because no scope has been assigned: in the scoped representation the handel will be the handel of the widest scoped quantifier. Scoped quantifiers have to be represented as having both a restriction and a body. However, for the underspecified representation the body of the quantifiers is left unspecified. The representation can be monotonically enriched to either scoped structure by appropriate coindexation of the handels and instantiation of the BODYs of the quantifiers. One of the scoped representations (wide scope *some*) is shown in Figure 12. Further details are given in Copestake *et al* (1995).

From now on, rather than simply using boxed integers to indicate reentrancy in the conventional way, we will indicate their types for clarity, using h for handel, e for event, x for entity, and i for

$$
\begin{bmatrix}
\text{HANDEL} & \boxed{5} \\
\text{INDEX} & \boxed{8} \\
\text{LISZT} & \left\langle
{}_{\textit{every\_rel}}\begin{bmatrix} \text{HANDEL} & \boxed{1} \\ \text{BV} & \boxed{2} \\ \text{RESTRICTION} & \boxed{3} \\ \text{BODY} & \boxed{4} \end{bmatrix},
{}_{\textit{dog\_rel}}\begin{bmatrix} \text{HANDEL} & \boxed{3} \\ \text{INST} & \boxed{2} \end{bmatrix},
{}_{\textit{some\_rel}}\begin{bmatrix} \text{HANDEL} & \boxed{5} \\ \text{BV} & \boxed{6} \\ \text{RESTRICTION} & \boxed{7} \\ \text{BODY} & \boxed{1} \end{bmatrix},
{}_{\textit{cat\_rel}}\begin{bmatrix} \text{HANDEL} & \boxed{7} \\ \text{INST} & \boxed{6} \end{bmatrix},
{}_{\textit{chase\_rel}}\begin{bmatrix} \text{HANDEL} & \boxed{4} \\ \text{EVENT} & \boxed{8} \\ \text{ACT} & \boxed{2} \\ \text{UND} & \boxed{6} \end{bmatrix}
\right\rangle
\end{bmatrix}
$$

Figure 12: MRS for *every dog chased some cat* with *some* taking wide scope

$$
\begin{bmatrix}
\text{HANDEL} & \boxed{h1} \\
\text{INDEX} & \boxed{e8} \\
\text{LISZT} & \left\langle
{}_{\textit{good\_rel}}\begin{bmatrix} \text{HANDEL} & \boxed{h4} \\ \text{ARG} & \boxed{e8} \end{bmatrix},
{}_{\textit{neg\_rel}}\begin{bmatrix} \text{HANDEL} & \boxed{h3} \\ \text{ARG} & \boxed{h4} \end{bmatrix},
{}_{\textit{suit\_rel}}\begin{bmatrix} \text{HANDEL} & \boxed{h5} \\ \text{EVENT} & \boxed{e8} \\ \text{ACT} & \boxed{i7} \\ \text{UND} & \boxed{x6} \end{bmatrix},
{}_{\textit{ego\_rel}}\begin{bmatrix} \text{HANDEL} & \boxed{h9} \\ \text{INST} & \boxed{x6} \end{bmatrix},
{}_{\textit{deixis\_rel}}\begin{bmatrix} \text{HANDEL} & \boxed{h10} \\ \text{ARG} & \boxed{i7} \end{bmatrix},
{}_{\textit{real\_rel}}\begin{bmatrix} \text{HANDEL} & \boxed{h2} \\ \text{ARG} & \boxed{h3} \end{bmatrix}
\right\rangle
\end{bmatrix}
$$

Figure 13: MRS for *That really doesn't suit me well*

individual (event or entity). For convenience, arguments to rels are shown as atomic entities in this document, but indices actually have a complex structure as in Pollard and Sag (1994). The MRS representation of (27) is shown in Figure 13.

(27) That really doesn't suit me well

Here the handels of rels which are not scoped with respect to one another are unified: in general whenever two *mrs-struct*s are combined, the handels are unified to create the handel for the result unless a rel in the liszt of one *mrs-struct* takes the handel of the other as an argument. Because scope has not been assigned, the outermost handel is not coindexed with any of the handels internal to the rels — it should be thought of as a disjunction of h2 and h10. The new liszt is constructed by appending the liszts of the structures.

The inclusion of the unresolved *ego_rel* and *deixis_rel* reflects the idea that the semantics should preserve enough surface information to facilitate translation for cases where there is a straightforward mapping between languages. This is not built into MRS in any way, however, so the representation is more flexible than QLF or Shake-and-Bake. Note that *well* has *good_rel* in its semantics and that *really* contains a *real_rel*. For adverbs where there is a systematic distinction in meaning from the corresponding adjective it would be necessary to use a separate rel. The current treatment does not preclude *suit*, *real* and so on from being polysemous in that there could be subsorts of *suit_rel*, *real_rel* and so on. As before we omit any representation of tense. Note that both *real_rel* and *neg_rel* take handels as arguments to allow for the difference between (28a) and (28b):

(28) a That really doesn't suit me

b That doesn't really suit me

This contrasts with PPs such as *on Monday* and *in Foothill Park* which are represented as taking events to avoid spurious ambiguity.

One very general assumption is that this approach is strongly but not completely lexicalist — the vast majority of rels will be introduced lexically (i.e. in lexical entries or lexical rules). The main exception to this are relations indicating the type of speech act, *y_n_q_rel*, *imp_rel* etc which are introduced syncategorematically. There will normally be a distinct rel for each sense of a word. Rel names usually indicate the orthography of the corresponding lexical item (e.g. *would_rel*). Senses are distinguished by words, not numbers. For example, *work_acceptable_rel* corresponds to the sense of *work* in (29):

(29) Does that work for you too?

Rels are not distinguished purely on the basis of syntax: thus *eat_rel* would cover both transitive and intransitive forms, the latter having an understood undergoer. Some rels are not of this form: for example *ego_rel* corresponds to *I* or *me*, *deixis_rel* to *that*, *neg_rel* indicates negation and *y_n_q_rel* and *imp_rel* stand for yes-no-question and imperative respectively.

Because rels are sorts, they will be in a hierarchy. This can be utilised in order to make the appropriate linguistic generalisations: for example, *come_move_rel* could be a subsort of *move_rel*, which would, among other things, allow a general expression of the alternations that apply to movement verbs. Note that this hierarchy will not usually correspond directly to the domain model concept hierarchy, since the latter is motivated by the real world rather than linguistic considerations.

### 6.2 Translation

Translation between MRS representations depends on setting up the correct reentrancies between the arguments of the rels, as with the previous systems we discussed.[8] As before, we initially show the example of translation of *das ist schlecht*, using a simplified semantics (Figure 14). As with tlinks, transfer equivalences can themselves be seen as feature structures: reentrancies are used between the two halves of the equivalences to indicate argument identification and generalizations about classes of transfer equivalence are specified using sorts. Note that the output MRS is underspecified, in that the top-level HANDEL value is not instantiated, but constraints on the English grammar would mean that h1 is the only possible value.

Most pure semantic transfer systems require *structural* transfer rules which handle the reordering and rearrangement of the semantic structures. Structural transfer rules are minimized by the MRS representation, which is desirable not only because of the reduction in number and complexity of the transfer rules, but because it is extremely difficult to guarantee that phrasal transfer rules and structural rules will interact properly. Transfer between two MRS representations essentially requires a single structural transfer rule, if we think of the value of LISZT as a set of rels:

[8] For simplicity, we show coindexation between the complete indices here, but the representation has to be slightly more complex, since in HPSG indices contain information which will not be shared between languages, such as gender.

Input sentence: Das ist schlecht

Input MRS:

$$\begin{bmatrix} \text{HANDEL} & \boxed{h1} \\ \text{LISZT} & \left\langle {}_{\mathit{deixis\_rel}}\begin{bmatrix} \text{HANDEL} & \boxed{h1} \\ \text{INST} & \boxed{x2} \end{bmatrix}, {}_{\mathit{schlecht\_rel}}\begin{bmatrix} \text{HANDEL} & \boxed{h1} \\ \text{ARG} & \boxed{x2} \end{bmatrix} \right\rangle \end{bmatrix}$$

Transfer equivalences:

$${}_{\mathit{deixis\_rel}}\begin{bmatrix} \text{HANDEL} & \boxed{h2} \\ \text{INST} & \boxed{x1} \end{bmatrix} \Longleftrightarrow {}_{\mathit{deixis\_rel}}\begin{bmatrix} \text{HANDEL} & \boxed{h2} \\ \text{INST} & \boxed{x1} \end{bmatrix}$$

$${}_{\mathit{schlecht\_rel}}\begin{bmatrix} \text{HANDEL} & \boxed{h3} \\ \text{ARG} & \boxed{x4} \end{bmatrix} \Longleftrightarrow {}_{\mathit{bad\_rel}}\begin{bmatrix} \text{HANDEL} & \boxed{h3} \\ \text{ARG} & \boxed{x4} \end{bmatrix}$$

Instantiated equivalences:

$${}_{\mathit{deixis\_rel}}\begin{bmatrix} \text{HANDEL} & \boxed{h1} \\ \text{INST} & \boxed{x2} \end{bmatrix} \Longleftrightarrow {}_{\mathit{deixis\_rel}}\begin{bmatrix} \text{HANDEL} & \boxed{h1} \\ \text{INST} & \boxed{x2} \end{bmatrix}$$

$${}_{\mathit{schlecht\_rel}}\begin{bmatrix} \text{HANDEL} & \boxed{h1} \\ \text{ARG} & \boxed{x2} \end{bmatrix} \Longleftrightarrow {}_{\mathit{bad\_rel}}\begin{bmatrix} \text{HANDEL} & \boxed{h1} \\ \text{ARG} & \boxed{x2} \end{bmatrix}$$

Output MRS:

$$\begin{bmatrix} \text{HANDEL} & \boxed{h1} \\ \text{LISZT} & \left\langle {}_{\mathit{deixis\_rel}}\begin{bmatrix} \text{HANDEL} & \boxed{h1} \\ \text{INST} & \boxed{x2} \end{bmatrix}, {}_{\mathit{bad\_rel}}\begin{bmatrix} \text{HANDEL} & \boxed{h1} \\ \text{ARG} & \boxed{x2} \end{bmatrix} \right\rangle \end{bmatrix}$$

Output sentence: That is bad

Figure 14: Transfer using MRS

$$\begin{bmatrix} \text{HANDEL} & \boxed{h1} \\ \text{INDEX} & \boxed{e8} \\ \text{LISZT} & \left\langle {}_{schlecht\_rel}\begin{bmatrix}\text{HANDEL} & \boxed{h3}\\ \text{ARG} & \boxed{e8}\end{bmatrix}, {}_{passen\_rel}\begin{bmatrix}\text{HANDEL} & \boxed{h5}\\ \text{EVENT} & \boxed{e8}\\ \text{THEME} & \boxed{i7}\end{bmatrix}, {}_{bei\_rel}\begin{bmatrix}\text{HANDEL} & \boxed{h5}\\ \text{ARG1} & \boxed{e8}\\ \text{ARG2} & \boxed{x6}\end{bmatrix}, \right. \\ & \left. {}_{ego\_rel}\begin{bmatrix}\text{HANDEL} & \boxed{h9}\\ \text{INST} & \boxed{x6}\end{bmatrix}, {}_{deixis\_rel}\begin{bmatrix}\text{HANDEL} & \boxed{h10}\\ \text{INST} & \boxed{i7}\end{bmatrix}, {}_{echt\_rel}\begin{bmatrix}\text{HANDEL} & \boxed{h2}\\ \text{ARG} & \boxed{h3}\end{bmatrix} \right\rangle \end{bmatrix}$$

Figure 15: MRS for *Das paßt echt schlecht bei mir*

**Definition 1 (Structural transfer rule)** *The translation $\tau$ of a set of rels $z = x \cup y$ is defined as*

$$\tau(x \cup y) = \tau_l(x) \cup \tau(y)$$

*where $\tau_l$ is the base case: $x$ translates as $\tau_l(x)$ iff there is some transfer equivalence such that the input unifies with $x$ giving $\tau_l(x)$ as the output.*

In other words, the translation of a *mrs-struct* is the translation of some subset of the liszt unioned with the translation of the remainder of the liszt.[9]

## 6.3 Transfer equivalences

All transfer equivalences consist of relationships between sets of rels but the vast majority will be lexically motivated. Some rels such as *neg_rel*, *ego_rel* and *deixis_rel* are not language-specific: in general we allow relations with a sort which is defined to be interlingual to simply be transferred as they stand between source and target. In general, however, transfer relationships have to be explicitly stipulated. For example, below we give examples of transfer equivalences that could apply in the translation of (30a) (assuming it has the MRS shown in Figure 15) to (30b):

(30) a das paßt echt schlecht bei mir

b that really doesn't suit me well

$$\left\langle {}_{echt\_rel}\begin{bmatrix}\text{ARG} & \boxed{h1}\end{bmatrix} \right\rangle \Longleftrightarrow \left\langle {}_{real\_rel}\begin{bmatrix}\text{ARG} & \boxed{h1}\end{bmatrix} \right\rangle$$

$$\left\langle {}_{passen\_rel}\begin{bmatrix}\text{EVENT} & \boxed{e3}\\ \text{THEME} & \boxed{i1}\end{bmatrix}, {}_{bei\_rel}\begin{bmatrix}\text{ARG1} & \boxed{e3}\\ \text{ARG2} & \boxed{x2}\end{bmatrix} \right\rangle \Longleftrightarrow \left\langle {}_{suit\_rel}\begin{bmatrix}\text{ACT} & \boxed{i1}\\ \text{UND} & \boxed{x2}\end{bmatrix} \right\rangle$$

$$\left\langle {}_{schlecht\_rel}\begin{bmatrix}\text{ARG} & \boxed{e1}\ event\end{bmatrix} \right\rangle \Longleftrightarrow \left\langle {}_{neg\_rel}\begin{bmatrix}\text{ARG} & \boxed{h2}\end{bmatrix}, {}_{good\_rel}\begin{bmatrix}\text{HANDEL} & \boxed{h2}\\ \text{ARG} & \boxed{e1}\end{bmatrix} \right\rangle$$

[9] Although this rule could be regarded simply as a constraint between input and output structures, it has to be implemented specially within constraint based formalisms which lack a representation of sets.

Note that we have refrained from relating the non-argument handels and event variables explicitly here. It is redundant to repeat the information that the handels and events are in one-to-one relationship for each rel, since it could simply be inherited from a general sort for transfer equivalences (in a manner comparable to the use of tlinks). The rule would apply to both adjectival and adverbial uses of *schlecht* (we assume that *good* and *well* both have *good_rel* in their semantics). However, we have specialized the antecedent so that it only applies when an event is modified.

Note that the use of handels allows a transfer relationship to be expressed by relating variables in a way which would be impossible with a more conventional semantic representation without adopting additional devices (such as transfer variables). In many cases, Shake-and-Bake transfer rules could be generated automatically from MRS rules by unifying the MRS rels with their corresponding lexical entries. MRS transfer is more flexible in that it does not require that there be corresponding lexical entries, thus it will work with grammar rules that introduce semantic information, such as *y_n_q_rel*. Expressing the relationship between rels rather than signs makes a more elegant treatment of some cases of translation mismatch possible: see the comparison in Copestake *et al* 1995. That paper also discusses how the mechanisms for generalisation which were developed for tlinks in ACQUILEX have straightforward counterparts within MRS.

### 6.4 Summary

**Coverage** At present, there are no obvious limitations, except those which are inherent to MRS itself (lack of a general treatment of higher-order constructions). This is irrelevant if the underlying representation is assumed to be DRT.

**Modularity** There is no explicit reference to the syntactic structure of the source or target languages. The logical form equivalence problem is mitigated in a way comparable to that in Shake-and-Bake.

**Independence** The use of partially specified transfer rules discussed in Copestake *et al* (1995) would make complex interactions possible, but otherwise rules are independent of one another.

**Underspecification** Various types of underspecification are possible in MRS and could straightforwardly be maintained in transfer.

**Generalisation** There are similar possibilities to those for tlinks. This is discussed in Copestake *et al* (1995).

**Interaction with generation** Transfer produces a logical form, which could be underspecified, requiring the generation component to make some lexical choices etc. It would also be possible for a strategic generation component to rearrange the logical form.

**Interaction with domain model etc** In principle, an MRS based approach can make use of the same sort of interaction as the MDS, and is more flexible in this respect than Shake-and-Bake. See the remarks in the next section.

**Implementation** The transfer process itself can be implemented in a rather similar way to the transfer between bags of signs used in Shake-and-Bake. Partially specified transfer rules would be more difficult to implement efficiently and correctly.

**Acquisition** MRS links are more abstract than tlinks and might be more difficult to write because of that, especially if the naming conventions for rels are unintuitive. But automatic acquisition techniques developed for tlinks should work equally well for MRS links.

# 7 Inference and interaction with generation

In the previous sections, we have discussed examples of translation equivalence which were constrained by parts of the logical form. However, many translation choices require more general inference mechanisms, which have to be implemented externally to the transfer mechanisms outlined above. These might involve relatively open ended reasoning using a dialogue model or a domain model, in contrast with the more limited possibilities for inference available within a typed unification system. We are not concerned here with exactly how such inference should be implemented but only with the interaction with transfer. In principle there are at least three options:

1. The transfer equivalence statement could require information from the domain/dialogue model.

   The *Termin* example given in §4 illustrates this approach.

2. Transfer could overgenerate, leaving generation to perform a semantic well-formedness check on the target LF to exclude those which are contextually impossible or implausible. This alternative is not attractive with a straightforward version of the lexicalist approach, because the full LF for the target is not produced prior to generation. The filtering mechanism would thus have to operate after generating the complete output sentence, and since this is the most computationally expensive step this is not an attractive option.

3. Transfer could generate a semantically correct target LF, which was reformulated by a strategic generation component in order to improve the output sentence. In general such reformulation will involve some inference. Similar remarks about the unsuitability of this approach for the lexicalist mechanism apply as above.

The first two options allow selection between translations — the third allows reformulation of a semantically plausible LF. There is also the possibility that transfer would overgenerate but that the target grammar would filter out some possibilities. This might be appropriate to block the generation of ?*that suits me badly*, for example, as we will discuss in §7.3. This is not a case of inference in the same way as the other options listed above, but we mention it here for completeness.

## 7.1 Inference for choice of translation equivalence

The most straightforward examples of a requirement for inference is where there are multiple possible translations of a word which cannot be distinguished by immediate context, but only by reasoning about the state of the dialogue etc. For example, as we mentioned in §4, *Termin* in German can be translated as *date* or *appointment* in the Verbmobil domain.[10] The following examples illustrate this from the Verbmobil dialogues:

(31) a Das ist schlecht - da habe ich um 14.30 einen Termin beim Zahnarzt.
That's bad - I have a dentist's appointment at 14:30.

b Ja ich habe schon davon gehört und habe den Termin noch freigehalten.
Yes, I've already heard about that and have reserved that date.

There appears to be a fundamental distinction between the *date* and *appointment* usages — the first is describing a slot and the second a filler. Getting this wrong in translation will normally be a serious error. For example:

[10] The situation is actually more complex than this, since neither translation is good in some contexts, where *meeting* would be preferable. We discuss this further in §7.3.

(32) That's bad - I have a date at the dentist at 14:30.

The difficulty here is partly that, although English also has both usages for *date*, it is restricted in the filler sense to social events, usually romantic or sexual encounters. But although there may sometimes be local clues about the translation, the reasoning involved may be quite complex and depend on a knowledge of the context of the source sentence. This is a very clear example where the domain model should be used to resolve the translation, because it allows the system to take advantage of the restricted topics of the dialogue.

Neither the tlink nor the MRS approach to transfer have mechanisms for accessing the domain model within the formalism. It would, of course, be possible to simply add an arbitrary condition statement to the expression of translation equivalences in either formalism.

$$_{\mathit{termin\_rel}}\begin{bmatrix}\text{INST} & \boxed{x1}\end{bmatrix} \Longrightarrow {}_{\mathit{date\_rel}}\begin{bmatrix}\text{INST} & \boxed{x1}\end{bmatrix}$$

Condition: veranstaltung_c (x1)

The alternative technique mentioned above would decouple the two mechanisms, and treat the inference mechanism as checking the output of transfer for plausibility (possibly with reference to the domain). For example, instead of the condition given above on the MRS form, there could be semantic well-formedness conditions which were applied to the English logical form, which would exclude the translation of *Termin* as *date*. In the long run this technique might be preferable (though more expensive if implemented naively) because it is not linked to particular transfer rules. However the task still should be seen as the rather directed one of distinguishing between options for translation, because this makes it constrained enough to implement. It is not plausible that a fully general model of plausibility of a logical form could be used, even for the restricted domain assumed for Verbmobil. This alternative is not attractive with a straightforward version of the lexicalist approach, because the full LF for the target is not produced prior to generation.

### 7.2 Inference for alternative expression

In contrast, consider the following example:

(33) a Können Sie noch einen anderen Terminvorschlag machen?
b Could you make another suggestion for a date?
c Could you suggest another date?

Here (33b) is more literal, but (33c) might be preferred: (33b) is not wrong, but it is rather lumpy. But if it is desirable to generate (33c) instead of (33b) some inference is required to achieve this. In the original sentence, the quantifier applies to *Vorschlag* (*suggestion*) rather than *Termin*. It would usually be the case that *einen anderen Terminvorschlag* implies that another date (or time) was proposed, but it is not generally true that a quantifier can be 'shifted' in transfer so that it applies to the predicate corresponding to the first half of a compound rather than the second: the validity of this move depends on the meanings of the compound and the determiner. For example, it is not correct to monotonically infer *suggest three dates* from *make three suggestions for a date*, as may be illustrated by the following:

(34) Kim has made three suggestions for a date: the 13th of November, the 13th of November and the 13th of November!

Note that it is plausible to assume that the LF corresponding to (33b) will actually be generated by the transfer mechanism unless it is somehow blocked. It is necessary to have a translation of

*Terminvorschlag* independently of *machen* because the sentence could be, for example:

(35) Diesen Terminvorschlag haben wir doch schon letzte Woche diskutiert.
We discussed that proposal for the date last week.

Similarly it must to possible to translate *machen* as *make* when it occurs with an object *Vorschlag*. Thus a simple combination of translation rules which will have to be provided for other sentences will naturally result in (33b). [11] Therefore there are potentially two ways of generating (33c): the LF corresponding to it could either be constructed as an alternative to (33b) by transfer from the German source, or the generation mechanism could produce it from the LF corresponding to (33b).

In the absence of a generator with a strategic component, the only option is for the transfer component to produce both possible LFs, and to impose conditions which will select between them. It is difficult to see how a strategic generation component could be integrated with a lexicalist approach to transfer. The same problem arises as that mentioned above with respect to checking the target LF for semantic plausibility: the full LF is only generated by the lexicalist mechanism as part of the operation of generating the target sentence — no LF is output by transfer component. Thus strategic generation does not combine well with the lexicalist approach.

This particular example illustrates some potential advantages of allowing manipulation of the target LF. As we mentioned in the introduction, the aim is to produce an architecture where transfer deals with differences between languages and not with monolingual issues. It is quite plausible that a generator for a monolingual system such as a natural language interface (NLI) could be given the LF appropriate for (33b), and that the same issue of possible conversion to (33c) would arise. In contrast, it is not likely that an NLI would produce an LF which contained a predicate which could either be realised as *date* or *meeting* — this is specific to the translation problem. Thus there seems to be some motivation for treating the output of (33c) rather than (33b) as a generation rather than a transfer problem. Another, more theoretical way of looking at this, is that preference for (33c) over (33b) could be seen as exemplifying the Gricean Maxim of Manner, which is a monolingual principle. In contrast, preference for (36b) as opposed to (36c) as a translation of (36a) is straightforwardly to do with the content of the utterance.

(36) a Das ist schlecht - da habe ich um 14.30 einen Termin beim Zahnarzt.
b That's bad - I have a dentist's appointment at 14:30.
c That's bad - I have a date at the dentist at 14:30.

Our conclusion is that allowing a strategic generation component to manipulate the output of transfer could have considerable advantages in simplifying the transfer component, improving the naturalness of the output and in generally increasing the robustness of the overall system. If this view is correct, the lexicalist approach to MT should be rejected. We assume that the generator can reformulate an input LF into a (near) logically equivalent form that can be naturally expressed in the target language without involving transfer. This sort of capability is necessary for generators in monolingual systems, so the proposal involves making use of existing techniques, in contrast to the more radical approach of translation by negotiation. For the example considered above, it was

[11] Transfer cannot produce an output which is completely neutral between these representations without loss of information and complication of transfer rules. The noun *suggestion* should have some part of its semantics in common with the verb *suggest* — e.g. both could contain a *suggest_rel*. This is necessary for purely monolingual reasons if *suggestion* is to be treated as derived from *suggest* by lexical rule. But the representation must also contain information about the scope of the quantifier, and to make this neutral between the translations of the two parts of the compound would be incorrect in general. Of course, this particular case could be treated as exceptional, and special transfer rules written to accomodate it. But this sort of approach leads to an inflexible system which cannot be scaled up.

clearly preferable to assume that the generator can adjust the LF rather than to complicate the translation rules so that transfer produces an underspecified LF. We would expect this to be true in general, since producing an underspecified form implicitly requires that the transfer rule writer has to consider all the LFs which the strategic generator might produce, which considerably reduces the advantages of having such a component.

### 7.3 Control

Explicit conditions stated within transfer rules are appropriate techniques for controlling the application of transfer rules in many circumstances. If the job of a semantic transfer rule is to ensure that denotations are (approximately) preserved, then the conditions will be restrictions on meaning of the source or target language predicates, but these may have to be stated in a variety of ways, since it is impossible to get at the intended meaning directly. The use of the domain terms mentioned in §4 is a relatively direct way to get at intended meaning, which should make it possible to resolve transfer ambiguities in some cases which analysis of the sentence cannot resolve without reference to the domain. In some circumstances this might be quite difficult — *Termin* as a filler can be translated as *meeting* or *appointment*. This is a case where there is a mismatch in meaning: *Termin* in this use intersects in meaning with both *meeting* and *appointment*. But the distinction between *appointment* and *meeting* depends on social factors, such as the relative status of the participants. Working out which was intended might involve an arbitrary amount of reasoning on real world knowledge, and a better strategy might be to use a statistical method, based on word cooccurrences.

In §2.3, when discussing the example of the translation of *geben*, we briefly mentioned the idea of interpreting transfer rules as defeasible implications. In this case, it would be natural to assume that a more specific transfer rule was applied in preference to a more general one. This idea was tried in ELU (Estival *et al*,1990). If condition statements are explicit, the specificity would presumably be calculated with reference to the conditions, rather than to the rule as a whole. It would probably be more appropriate to formalise this within a logic which is credulous, rather than skeptical: that is, if two rules both have their antecedents verified and are not in a mutual specificity relationship, then both rules should apply.

However, there are cases where it is not appropriate to put explicit conditions on transfer rules. For example, the translation of *schlecht* as *badly* is unavailable when translating (37a) not because there is anything wrong with the presumed meaning of (37b) but because there is a monolingual restriction in English that *suit*, along with other 'positive' psychological predicates, cannot take an adverbial denoting a negative attitude.

(37) a Das paßt schlecht bei mir.
b * That suits me badly.
c That doesn't suit me at all.

The transfer component has to provide an alternative, but it is a monolingual condition that blocks the usual translation. (37b) is not (fully) acceptable in any context. Contrast this with *geben* vs *give*: there is nothing *linguistically* wrong with saying *you gave me that book*, but it means something different from the German sentence *du gibst mir das Buch*, in some contexts. Similarly, as we mentioned briefly in §3.1, it is not the responsibility of a semantic transfer component to deal with all the idiosyncrasies of English preposition selection. In these cases, it is better to rely on the generator/target grammar filtering the proposals made by transfer. This is a very limited step toward the translation by negotiation proposal, but unless there is a (relatively) clear demarcation of this sort, it will not be possible to have any sort of principled approach to transfer.

In principle, therefore, syntactic conditions on semantic transfer should be unnecessary. If there is a difference in meaning between two syntactically distinct usages of a word, this should be reflected in the semantics. If there is not, the distinction should be irrelevant for semantic transfer. Syntactic conditions indicate that the monolingual semantic analysis is wrong, or insufficient. As a temporary expedient, a semantic transfer component may have to allow syntactic conditions, but this indicates a flaw in the monolingual analysis.

## 8 Conclusion

Part of the aim of this paper has been to try and present a view of semantic transfer as a subject in its own right. We have argued that, for the foreseeable future, semantic transfer will be the best approach to MT for systems such as Verbmobil, because of the limitations on practical discourse processing and inference. We have suggested that some of the problems often assigned to transfer are actually problems for the monolingual target grammar and generation component. The hope is that, if we can get the interplay between transfer and monolingual components right, transfer rules will be less arbitrary, and it will be easier to write and maintain them.

The three approaches described in detail in this paper have a considerable amount in common, with respect to reliance on constraints/unification in a typed framework, specification of equivalence at a semantic level, modularity, (relative) independence of application of transfer rules, ability to cope with partial interlingua, and (some degree of) lexicalist orientation. All these features have significant utility.

The MDS transfer component works efficiently on a range of Verbmobil dialogues, and makes good use of the constructs developed for semantic construction. It has certain inelegancies, some of which are due to the restrictions put on its design by the other components in the Verbmobil demonstrator system. In particular, this approach would probably work better with a 'flatter' semantic representation, such as MRS. It is clear that some hacks were necessary to get it to work for the examples, and this limits confidence in how well it would scale up without modification. Minimally, it appears that some redesign is needed to cope with phrasal translations. We suggest that it would be useful in the long term to attempt to separate the declarative statement of transfer equivalences from the control of their application.

Lexicalist machine translation is an attractive approach, which has a number of advantages and which has been very influential in recent work on MT. An efficient generation algorithm for Shake-and-Bake has now been developed (Poznanski *et al* 1995), which removes one of the obstacles to its adoption. However the approach is unable to cover some phenomena without extension and will not work (in its current form) with grammars which are not completely lexicalist. For long-term development, its inability to interact with full flexibility with an inference and strategic generation component is a serious disadvantage.

As we have described here, MRS is primarily an approach to semantic representation rather than to transfer. We think that adopting it (or another approach to flat underspecified semantic representation) will considerably improve the prospects for writing maintainable and extensible transfer rulesets. The specific approach to transfer suggested in the context of the work on MRS was an attempt to take the ideas in Shake-and-Bake and to apply them to pure semantic transfer. From a theoretical viewpoint, it appears that this would go some way towards correcting the deficiencies noted in Shake-and-Bake, while maintaining many of the advantages of that method. For the MRS approach to transfer to be realistic however, it would have to adopt some approach to specification of conditions on transfer rules. It is also necessary to determine how well MRS itself works with generation systems and whether the aim of distinguishing between the declarative statement of transfer equivalences and control principles can be made to work in practice.


## Acknowledgements

I am grateful to Martin Emele and Dan Flickinger for their help in preparing this report, and especially to Susanne Riehemann for discussion of the German examples. However, all opinions and mistakes are mine.


## References


Alshawi, H., D. Carter, M. Rayner and B. Gambäck (1991) 'Translation by quasi logical form transfer', *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics (ACL-91),* Berkeley, California, pp. 161–168.

Apresjan, Ju D. (1973) *Regular Polysemy,* Mouton, The Hague, The Netherlands.

Arnold, D. and L. Sadler (1990) 'The theoretical basis of MiMo', *Machine Translation,* **5 (3)**, 195–222.

Beaven, J. L. (1992) 'Lexicalist unification-based machine translation', PhD thesis, University of Edinburgh.

Brown, P. F., J. Cocke, S.A. DellaPietra, V.J. DellaPietra, F. Jelinek, J.D. Lafferty, R.L. Mercer and P.S. Roosin (1990) 'A statistical approach to Machine Translation', *Computational Linguistics,* **16(2)**, 79–85.

Butt, M. (1995) 'Transfer I: Tense and Aspect', Verbmobil Report 55, Eberhard-Karls-Universität Tübingen.

Copestake, A. (1993) 'The Compleat LKB', Technical report 316, University of Cambridge Computer Laboratory.

Copestake, A. and A. Sanfilippo (1993) 'Multilingual lexical representation', *Proceedings of the AAAI Spring Symposium on Building lexicons for machine translation,* Stanford, CA.

Copestake, A., E.J. Briscoe, P. Vossen, A. Ageno, I. Castellon, F. Ribas, G. Rigau, H. Rodriguez, A. Samiotou (1995) 'Acquisition of lexical translation relations from MRDs', *Machine Translation,* **9(3)**.

Copestake, A. and E.J. Briscoe (1995) 'Semi-productive polysemy and sense extension', *Journal of Semantics,* **12:1**, 15–67.

Copestake, A., D. Flickinger, R. Malouf, S. Riehemann, I. Sag (1995) 'Translation using Minimal Recursion Semantics', *Proceedings of the The Sixth International Conference on Theoretical and Methodological Issues in Machine Translation (TMI95),* Leuven, Belgium.

Davis, A. (forthcoming) 'Linking and the hierarchical lexicon', Doctoral dissertation, Stanford University.

Dorr, B. J. (1994) 'Machine Translation divergences: a formal description and proposed solution', *Computational Linguistics,* **20(4)**, 597–633.

Dörre, J. and R. Seiffert (1991) 'Sorted feature terms and relational dependencies', IWBS Report 153, IWBS, IBM Deutschland.

Emele, M., U. Heid, S. Momma and R. Zajac (1992) 'Interactions between linguistic constraints: procedural vs. declarative approaches', *Machine Translation,* **7 (1–2)**, 61–98.

Estival, D., A. Ballim, G. Russell and S. Warwick (1990) 'A syntax and semantics for feature structure transfer', *Proceedings of the 3rd International Conference on theoretical and methodological issues in MT of NLs (TMI-90),* Austin,Texas, pp. 131–143.

Frank, A. and U. Reyle (1994) 'Principle based semantics for HPSG', to appear in: Arbeitspapiere des Sonderforschungsbereichs 340, University of Stuttgart.

Hutchins, W.J. and H.L. Somers (1992) *An introduction to Machine Translation,* Academic Press, London.

Kamp, H. and U. Reyle (1993) *From Discourse to Logic,* Kluwer Academic Publishers, Dordrecht, Germany.

Kaplan, R., K. Netter, J. Wedekind and A. Zaenen (1989) 'Translation by Structural Correspondences', *Proceedings of the 4th Conference of the European Chapter of the Association for Computational Linguistics (EACL-89),* Manchester, UK, pp. 272–281.

Kay, M., J-M. Gawron and P. Norvig (1994) *Verbmobil: A translation system for face-to-face dialog,* CSLI Lecture Notes, No. 33: University of Chicago Press.

Landsbergen, J. (1987) 'Isomorphic grammars and their use in the ROSETTA translation system' in M. King (ed.), *Machine Translation Today: The State of the Art,* Edinburgh University Press, pp. 351–372.

Nirenburg, S. and V. Raskin (1988) 'The Subworld Concept Lexicon and the Lexicon Management System', *Computational Linguistics,* **13(3-4)**, 276–289.

Pollard, C. and I. Sag (1994) *Head-driven phrase structure grammar,* Chicago University Press, Chicago.

Poznanski, V., J.L. Beaven and P. Whitelock (1995) 'An efficient generation algorithm for lexicalist MT', *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics (ACL-95),* Cambridge, Mass..

Sag, I. (1995) 'English relative clause constructions', ms. Stanford University.

Sanfilippo, A., E. J. Briscoe, A. Copestake, M. A. Marti and A. Alonge (1992) 'Translation equivalence and lexicalization in the ACQUILEX LKB', *Proceedings of the 4th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-92),* Montreal, Canada.

Shieber, S.M. (1993) 'The problem of logical form equivalence', *Computational Linguistics,* **19 (1)**, 179–190.

Slocum, J., W.S. Bennet, J. Bear, M. Morgan and R. Root (1987) 'METAL: The LRC Machine Translation system' in S. Michaelson and Y. Wilks (ed.), *Machine Translation today: the state of the art,* Edinburgh University Press, pp. 319–350.

Trujillo, I.A. (1995) 'Lexicalist Machine Translation of Spatial Prepositions', PhD dissertation, University of Cambridge.

Whitelock, P. (1992) 'Shake-and-bake translation', *Proceedings of the 14th International Conference on Computational Linguistics (COLING-92),* Nantes, France.

Zajac, R. (1989) 'A transfer model using a typed feature structure rewriting system with inheritance', *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics (ACL-89),* Vancouver, BC, pp. 1–7.